

-2299-

||| ||:|||||||:|:| ||||| ||||| |||: ||||| |: |||:||||:| || ||:|
SPQEIEALVKHLMSTAKAHVNKETTAYTSRQQAGIIDTAAAISTGLYLTG-EDGYGSTLGNVEDTFSFTVTLHNIT
610 620 630 640 650 660 670

5 2400 2430 2460 2490 2520 2550 2580 2610
KVAKDLHYTTYLNIDQVKDGFVTLAPQQLGTFGTGKTIRIEPGQTITITIDIDVSKYHMLKKVMPNGYFLEGYVRFTDPV
||:|:| || : | : : : : : |:|:| | : : | : |||:|:|:|
NEDKTLNYSTQLTTDTAQKRIDLHGSTSISRDSWRKVTVKANSSTTVTINVDASSFAEELTGLMKNGYYLEGFVRFTDVA
690 700 710 720 730 740 750

10 2640 2670 2700 2730 2754 2784 2814 2844
DGGEVLSIPYVGFKEFQNLLEVLEKSIYKLVANKEKGFYFQP--KQTNEVPGSEDTALMTTSSEPIYSTDGTSEIQLKA
| : : : |||||:| ||| ||:| : |||:| | | | : | | | | : |
DDGDIVSIPYVGFGRFQNLAVLEPIYNLIADGKGGFYFEPVTAQPNVTVDISHHYTGLVTGSTELIYSTDKRSDSAIKT
770 780 790 800 810 820 830

15 2874 2904 2934 2964 2994 3024 3054 3084
LGSYKSIDGKWILQLDQKGPHLAISPNDQNDQAVAVKGVFLRNFNNLRAKVYRADDVNLQKPLWVSAPOAGDKNYVSG
||:|:| : ||:|:| : ||||| |||:| : |||||:| : | | | | | ||| | | :|:|:|
LGTFFKNAGYFVLELDESGKPHLAISPENDDNQDSLVEKGVFLRNYTDLVASVYAADDTERTNPLWESQPSGDKNIYSG
850 860 870 880 890 900 910

20 3114 3144 3174 3204 3234 3264 3294 3324
NTENPKSTFLYDTEWKGTTTDDGIPLEDGKYKYLVITYSYSDVPGSKPQMVFDITLDRQAPTITATYDKDRRIKARPAVE
| : |||:| : ||| || : || ||||:| ||| ||| : |:|:| : ||:|:|:|:|:|:|:| | |||:|
NPKNPKSSIYPTEWNGTSDGNALADGKYQYVLTYSKVPGAQVQTMIFDVIIDRESPVITATYDETNTFTNFRPAIE
930 940 950 960 970 980 990

25 3354 3384 3414 3444 3474 3504 3534 3564
HGESGIFREQVFLKKDKDGHYNSVLROQGEDGILVEDNKVFIKQEKDGSFILPEKVDNDFSHVYTVEDYAGNLVSAKLE
|||:|:| ||| | : : : | ||||:| | ||| || : : | | |||||:|:| :|
KGESGLYREQVFLVADASG-VTTIPSLKNGDVTVSDNKVFVAQNDDGSFTLPLDLADISKFYTVEDYAGNISYKVE
1010 1020 1030 1040 1050 1060 1070

30 3594 3624 3654 3684 3711 3741 3771 3801
DLINIGNKNGLVNVKVPSPELNSNVDIDFSYSVKDDKGNIIKK-QHHGKDLNLLKLPFGTYTFDLFLYDEERANLISPKS
:|:|:|:| ||| | : : | | ||||| :| :| : : | :|:|:|:|:|:|:|:| | :|
NLISIGNEKGLVTNILDKDTNSPVPILFSYSVTDETKIVAELPRYAGDTSVLKLPFGTYTFDLFLYDTEWSSLAGETK
1080 1090 1100 1110 1120 1130 1140 1150

35 3831 3861 3891 3921 3951 3981 4011 4041
VTVTISEKDSLKDVLFKVNLLKKAALLVEFDKLLPKGATVQLVTKTNTVVDLPKATYSPTDYGKNIPVGVYRLNVTLPSG
||| | : | : | | | ||:| | ||| |:| ||| : || | | ||||| :||| | : ||| |
AVVTILEDNSTAEVNFVTLKDKANLLIDALLPSSGSTIQLVTDGQAIQLPNAKYSKTDYGFVPVGTYTILETLPEG
1160 1170 1180 1190 1200 1210 1220 1230

40 4071 4101 4131 4161 4191 4221 4251 4281
YSTLENLDDLLSVKEDQVNLTKLTLINKAPLINALAEQTDITQPVFYNAGTHLKNMYLANLEKAQTLIKNRVEQTSID
| | || | :| :| :| ||||| | :| :| :| :| | | | | :| :| :| :|
YEFLELD---VAVLANQSNVKKLTLINKVALKELIAELAGLEETARYYNASPELQTAYAKALEDANAVYANKHNQAQVD
1240 1250 1260 1270 1280 1290 1300

45 4311 4341 4371 4401 4431 4461 4491 4521
NAIAALRESQALNGKETDTSLLAKAILAETIKGNYQFVNASPLSQSTYINQVLAKNLLQKPNVTQSEVDKALENLDI
:|:|:| :|:|:|:| | : | : : : | | | :|:|:| :|:|:|:|:|:|:|:| :|:|:| :|
SALASLVAAREQLNGQATDKEKLIABVSNYTPTQANFIYYNAENTKQIAYDTAVRSAQLVLNQENVTQAVVNQALADLLA
1320 1330 1340 1350 1360 1370 1380

50 4551 4581 4611 4641 4671 4701 4731 4761
AKNQLNGHETDYSGLHHMIKANVLKQTSKYQNASQFAKENYNNLIKKAEELLSNRQATOQAVEELLNQIKATEQELDG
|| |:|:| | | : :||| | :| |||:| :| : :|:| :|:| :|:| :|:| :|:| :|:|
AKANLDGQKTDISALRAVSVSVLKATDAKYLNASENVKQAYDQAVEAAKAILVDESASQASVDQALAVLTSQAQELDG
1400 1410 1420 1430 1440 1450 1460

55 4779 4809 4839 4845 4860 4890 4920 4950
----RDRVSSAENYSQSLNDNDSLNTTPIN-----PP-----NQPALIFPKGMTKESEVAQKRVLVGTVTSQTDNQKV
: | : | : : ||: | | : | :| : : : : | : : ||:
VATSTNDAKEPANTATDKKDEGTVPPTPPIIDSEIVDVQAPPVKDTGNSEHVPIGQKPNPQPT-LPRPVTLQASLSSPENQEK

-2300-

1480 1490 1500 1510 1520 1530 1540

4980 5010 5040 5070 5100 5130 5160 5190

5 KTNKLPKGTGESTPKITVTYLLFSLMLGLATIKLKSIRE*NTLKNRARRHQLLAINS**LVPF*GA*NDVPKDLFSAVSW

: :||| :| :| :| :| :| :|

QVTQLPNTGENDTK--YYLVPGVIIGLGTLLVSIRRHKEEV

1560 1570 1580

10 A related GBS nucleic acid sequence <SEQ ID 10965> which encodes amino acid sequence <SEQ ID 10966> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6297> which encodes the amino acid sequence <SEQ ID 6298>. Analysis of this protein sequence reveals the following:

LPXTG motif: 1614-1619

15 Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -4.46 Transmembrane 1623 -1639 (1621 -1641)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.2784(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:AAG09771 GB:AF243528 cell envelope proteinase [Streptococcus thermophilus]
Identities = 465/1125 (41%), Positives = 668/1125 (59%), Gaps = 61/1125 (5%)

30 Query: 1 VEKKQRFSLRKYKSGTFSVLIGSVFLVM-TTTVAADELSTMSEPTITNHAQOQAQHLTNT 59
++KK+ FSLRKYK GT SVL+G+VFL +VAADEL+++ E + T
Sbjct: 1 MKKKEITFSLRKYKIGTVSVLLGAVFLFAGAPSVAADELSTLVETKVEA-----T 49

35 Query: 60 ELSSAESKSQDTSQITLKTNREKEQSQDLVSEPTTTELADTDAASMAN TGSDATQKSASL 119
+ S+S S + E+ D E T+T++ TD GS+A + SA
Sbjct: 50 VPD AIVSESASESPVV-----EELVDTSV EATSTDVTTDNEE-ETPGSEALENSA-- 99

40 Query: 120 PFVNTDVHDVWVKTKGAWDKGYKGQGVVAVIDTGIDPAHQSMRISDVSTAKVKSKE DMLA 179
NT+V T+ A + + KV + + ++D +TA +E
Sbjct: 100 ---NTEVET---TQPAVETPAISEKKV-----EEEEKLSVADETTAITNQEE---- 140

45 Query: 180 RQKAAGINYGSWINDKVFAHNVYENS DNIKE-NQFEDFDEWENFEFDAEAEPAIKKH 238
K I+ + I V+ Y + + D D D + + A+ K+ K+
Sbjct: 141 -AKPQNIDSNTIITVPKVWYSGYKGEGTVVAIDSGLDVDHDLHISDLSTAKYKSEKEI 199

50 Query: 239 KIYRPQSTQAPKETVIKTEETDGS HDIDWTQTD DDTKYESHGMHVTGIVAGNSKEAAATG 298
+ + + E + G + +D + SHGMHVT I GN + A G
Sbjct: 200 EAAKEAAGITYGEW-FNDKVVF GYNYVDVNTVLKEEDKRSHGMHVTSIATGNPTQPV A-G 257

55 Query: 299 ERFLGIAPEAQVMFMRVFPANDIMGSAESLFKAIEDAVALGADV INLSLGTANGAQLSGS 358
+ G+APEAQVMFMRVFP++ + +L++KAIEDAV LGAD INLSLG ANG+ ++ +
Sbjct: 258 QLMYGVAPEAQVMFMRVFSDLKATTGAALYVKAIEDAVKL GAD SINLSLGANGSVNMN 317

60 Query: 359 KPLMEAIEAKKAGVSVVVAAGNERVYGS DHDPLATNP DYGLVGSPTGRTPTSVAAIN 418
+ + AIE A++AGVSVV+AAGN+ +GS H +P A PDYGLVG+PST SVA+ N
Sbjct: 318 ENVTAAIEAARRAGVSVVIAAGNDGTFGSGHSNPSADYPDYGLVGAPSTAHD AISVASYN 377

Query: 419 SKWVIQRLMTVKELENRADLNHGKAIYSESVD FDKIDKSLGYDKSHQFAYVKESTDAGYN 478
+ V +++ + LEN ADLN+GK+ + + + +G + + +A + +++D ++
Sbjct: 378 NTTVGSKVINIIGLENNADLN YGKSSF-DNPEKSPVPFEIGKEYEYVYAGIGQASD--FD 434

Query: 479 AQDVKGKIALIERDPNKTYDEMIALAKKHGALGVLI FNNKPGQSNRSMRLTANGMGIPSA 538
D+ GK+ALI+R T+ E IA A GA+GV+IFN++PG++N SM+L + IPS
Sbjct: 435 GLDLTGK LALIKRG-TITFSEKIANATAAGAVGVVIFNSRPGEANVSMQLDDTAIAIPSV 493

-2301-

5 Query: 539 FISHEFGKAMSQNLNGNGTGSLEPDSVSVSKAPSQKGNEMNHFSNWGLTSDGYLKPDPITAPG 598
FI EFG+A++ + + F++ P+ + ++ FS+WGL++DG LKPD+ APG
Sbjct: 494 FIPLEFGEALAA---NSYKIAFNWNETDIRPNFEAGLLSDFSSWGLSADGELKPDLAAPG 549

10 Query: 599 GDIYSTYNDNHYSQTGTSMASPOIAGASLLVKQYLEKTQPNLPKEKIADIVKNLLMSNA 658
G IY+ NDN Y + GTSMASP +AGA++LVKQYL T P ++I +VK+LLMS A
Sbjct: 550 GAIYAAINDNDYANMQGTSMASPHVAGAAVLVKQYLLATYPTKSPQEIEALVKHLLMSTA 609

15 Query: 659 QIHVNPETKTTTSPRQQAGLLNIDGAVTSGLYVTGKDNYSISLGNITDTMTFDVTVHN 718
+ HVN ET TSPRQQAG+++ A+++GLY+TG+D YGSI+LGN+ DT +F VT+HN
Sbjct: 610 KAHVNKETITAYTSPRQQAGIIDTAAAIISTGLYLTGEDGYGSITLGNVEDTFSFTVTLHN 669

20 Query: 719 LSNKDKTLRYDTELLTDHVDPOKGRFTLTSHSLKTYQGGEVTVPANGKVTVRVMTDVSQF 778
++N+DKTL Y T+L TD + TS S +++ +VTV AN TV + +D S F
Sbjct: 670 ITNEDKTLNYSQTLLTDTAQKRIDLGLSTISRDSWR--KVTVKANSSTTVTINVDASSF 727

25 Query: 779 TKELTKQMPNGYYLEGFVFRFRDSQDDQLNRVNIPFVGFKGFENLAVAEESIYRLKSQGK 838
+ELT M NGYYLEGFVRF D DD + V+IP+VGF+G+F+NLAV EE IY L + GK
Sbjct: 728 AEELTGLMKNGYYLEGFVRFDTVADDG-DIVSIPYVGRGEFQNLAVLEETIYNLIADGK 786

30 Query: 839 TGFYFDE-SGPKDDIYVGKHFTGLVTLGSETNVSTKTISDNGLHTLGTfKNADGKFILEK 897
GFYF+ + + + H+TGLVT +E ST SD+ + TLGTfKN G F+LE
Sbjct: 787 GGFYFEPVTAQPNVTDISHHYTGLVGTSTELIYSTDKRSDSAIKTLGTfKNKAGYFVLEL 846

35 Query: 898 NAQGNPVLAIISPNGDNNQDFAAFKGVFLRKQGLKASVYHSDKEHKNPLWVS-PESFKG 956
+ G P LAISPNGD+NQD FKGVFLR Y L ASVY A D E NPLW S P+S G
Sbjct: 847 DESGKPHLAISPNGDDNQDSLIVFKGVFLRNYTDLVASVYAADDERTNPLWESQFQS--G 904

40 Query: 957 DKN-FNSDIRFAKSTTLTGTAFGSKSLTGAELPDGHYHYVVSYPDVVGAQRQEMTFDMI 1015
DKN ++ + + KS+ + T ++G G L DG Y YV++Y V GA Q M FD+I
Sbjct: 905 DKNYISGNPKNPKSSIIYPTWNGTDSGDNALADGKYQYVLTYSKVPGAQVMTIMFDVI 964

45 Query: 1016 LDRQKPVLSQATFDPETNRFPKPEPLKDRGLAGVRKDSVFYLERKDNKPYTVTINDSYKYV 1075
+DR+ PV++ AT+D F P P ++G +G+ ++ VFYL + T+ V
Sbjct: 965 IDRESFVITTATYDETNTFTNPRPAIEKGESGLYREQVFYLVADASGVTTIPSLKNGDV 1024

50 Query: 1076 SVEDNKTVERQADGSFILPLDKAKLGDFFYMVEDFAGNVAIAKL 1120
+V DNK FV + DGSF LPLD A + FYY VED+AGN++ K+
Sbjct: 1025 TVSDNKVFAQNDDGSFTLPLDLADISKFYTTVEDYAGNISYEKV 1069

An alignment of the GAS and GBS proteins is shown below.

Identities = 543/1676 (32%), Positives = 821/1676 (48%), Gaps = 158/1676 (9%)

45 Query: 24 KQRFSSIRKYKLGAIVSVLLGLTFLFGLGITNVAAD--SVINKPSDIAVEQQVKDSPTSIS--- 78
KQRFSSIRKYK G SVL+G++F + T VAAD S +++P+ QQ T+
Sbjct: 4 KQRFSSIRKYKSGTFSVLIGSVFLVM-TTVADELSTMSPTITNHAQQQAQHLNTELS 62

50 Query: 79 ANETPTNNTSSALASTAQD-----NLVTKANNSPTETQPVASHSQATETTFSPVANQPVE 133
+ E+ + +TS T ++ +LV++ + A + ++ A+ P
Sbjct: 63 SAESKSQDTSQITLKTNREKESQDLVSEPTTTELADTDAASMAN TGSDATQKSASLPPV 122

55 Query: 134 STQEVSKTPLTKQ--NLAVKSTPAISKETPQID-SNKIIITVPKVWNTGYKGECTVVAI- 189
+T +V TK + K + ID +++ + + V K + ++A
Sbjct: 123 NT-DVHDWVKTKGAWDKYKQGKVVAVIDTGIDPAHQSMRISDVSTAKVKSKEDMLARQ 181

60 Query: 190 ----IDSGLDIN-----HDALQLNDSTKAK-----YQNEQQMNAAKAKAGINYGKW 231
I+ G IN H+ ++ +D+ K ++N + A+ KA I K
Sbjct: 182 KAAGINYGSWINDKVVFHNVVENSNDNIKENQFEDFEDWENFEDAEAPKA-ICKHKI 240

65 Query: 232 YN-----NKVIFGHNYVDVNTLKEVKSTSHGMHVTSIATANPSKKD-TNEL 277
Y + G + +D + K SHGMHVT I N + T E
Sbjct: 241 YRPQSTQAPKETVIKTETDGDHIDWTQTDGDKYESHGMHVTGIVAGNSKEAATGER 300

Query: 278 IYGVAPAEQVMFMRVFSDEKRGTPALYVKAIEDAVKLGADSNLSLGGANGSLVNADDR 337
G+APEAQVMFMRV+++ G+ +L++KAIEDAV LGAD INLSLG ANG+ ++
Sbjct: 301 FLGIAPEAQVMFMRVVFANDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQLSGSKP 360

-2302-

Query: 338 LIKALEMARLAGVSVVIAAGNDGTFGSGASKPSALYPDYGLVGSPTAREAISVASYNNT 397
 L++A+E A+ AGVSVV+AAGN+ +GS P A PDYGLVGSPT R SVA+ N+
 Sbjct: 361 LMEAIEKAKKAGVSVVVAAGNERVYGSDDHDDPLATNPDYGLVGSPTGRTPTSVAAINSK 420

Query: 398 TLVNKVFNIIGLENNRNLNGLAAYA---DPKVSDKTFEVKGQYDYVFGKGNNDNDYKDK 454
 ++ ++ + LEN +LN+G A Y+ D K + K + + +V + D Y +
 Sbjct: 421 WVIQRLMTVKELNRLNGLAAYA---DPKVSDKTFEVKGQYDYVFGKGNNDNDYKDK 480

Query: 455 TLNGKIALIERG-DITFTKKVNVNAINHGAVGAIIFNNKAGEANLTMSLDPEASAIPIAIFT 513
 + GKIALIER + T+ + + A HGA+G +IFNNK G++N +M L IP+ F
 Sbjct: 481 DVKGKIALIERDENPKTYDEMIALAKKHGALGVLIFFNNKPGQSNRSMRLTANGMGIPSAPI 540

Query: 514 QKEFGDVLAKNNYK----IVFNNIKNQANPNAGVLSDFSSWGLTADGQIKPDLSPAGGS 569
 EFG +++ N + F+++ +K + ++ FS+WGLT+DG LKPD++APGG
 Sbjct: 541 SHEFGKAMSQNLNGNGTGSLEFDSVVSAPSQKGNEMNHFSNWGLTSDGYLKPDIAPGGD 600

Query: 570 IYAAINDNEYDMMSGTSMASPHVAGATALVKQYLLKEHPELKKGDIERTVKYLLMSTAKA 629
 IY+ NDN Y +GTSMASP +AGA+ LVKQYL K P L K I VK LLMS A+
 Sbjct: 601 IYSTYNDNHYSQGTGTSMASQIAGASLLVKQYLEKTQPNLEKEKIADIVKNLIMSNAQI 660

Query: 630 HLNKDTGAYTSPRQQGAGIIDVAAVQITGLYLTGGENNYGVSITLGNIKDKISFDVTVHNI 689
 H+N +T TSPRQQGAG++++ AV +GLY+TG ++NYGS++LGNI D ++FDVTVHN+
 Sbjct: 661 HVNPETKTITSPRQQGAGLLNIDGAVTSGLYVTG-KDNYGSISLGNITDITMTFDVTVHNL 719

Query: 690 NKVAKDLHYTTYLNTDQV--KDGFTLAPQQLGFTTGKTIRIEPGQTQITITIDIDVSKYH 747
 + K L Y T L TD V + G TL L T+ G + + T+ + +DVS++
 Sbjct: 720 SNKDKTLRYDTELLTDHVDPPQKGRFTLTSLSLKYQGGEVTVPANGKVTVRVTMDVSOFT 779

Query: 748 DMLKKVMPNGYFLEGYVRFTDPVDGG-EVLSIPYVGFKGEFQNLVLEKSIYKLVANKEK 806
 L K MPNGY+LEG+VRF D D ++IP+VGFKG+F+NL V E+SIY+L + +
 Sbjct: 780 KELTKQMPNGYFLEGYVRFRDSDQDLNRVNI PFVVGFKGQFENLAVAEEISYRLKSQKGT 839

Query: 807 GFYFQPK-QTNEVPGSEDTALMTTSSEPIYSTDGTSPQLKALGSYKSIDGKWILQLDQ 865
 GFYF +++ + +T L+T SE ST S L LG++K+ DGK+IL+ +
 Sbjct: 840 GFYFDESQPKDDIYVGKHTGLVTLGSETNVSTKTISDNLHLLTGFKNADGKFILEKNA 899

Query: 866 KGQPHLAISPNDQNDQDAVAVKGVFLRNFNNLRAKVYRADDVNLOKPLWVSAPQ-AGDKN 924
 +G P LAISP N D NQD A KGVFLR + L+A VY A D + PLWVS GDN
 Sbjct: 900 QGNFVLAISPNGDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKGDN 959

Query: 925 YYSNTENPKSTFLYDTBWKGTTTDTGIPLEDGKYKYVLTYYSDVPGSKPQMVFDITLDR 984
 + S + KST L T + G + G L DG Y YV++YY DV G+K Q+M FD+ LDR
 Sbjct: 960 FNS-DIRFAKSTTLTGAFSGKSLTGAEPLDGHYHYVVSYPDVVGAQRQEMTFDMILDR 1018

Query: 985 QAPTTLTATYDKDRRIKPARPAVEHGESGIFREQVFYLLKDKDGHYNSVLRQQGEDGILV 1044
 Q P L+ AT+D + FK P + G +G+ ++ VFYL++ KD +V + V
 Sbjct: 1019 QKPVLSQLATFDPEITNRFKPEPLKDRGLAGVRKDSVFYLER-KDNKPYTFTINDSYKYVSV 1077

Query: 1045 EDNKVFIKQEKDGSFILPKEVNDFSHVYYTVEDYAGNLVSAKLEDLINIGNKGLNVNVK 1104
 EDNK F++++ DGSFILP + YY VED+AGN+ AKL D + + +K+
 Sbjct: 1078 EDNKTFVERQADGSFILPLDKAKLGDFYVNVEDFAGNVIAIKLGDLHPQTLGKTPIKLKL 1137

Query: 1105 FSPBLNSNVDIDFSYSVKDDKGNIIKKQ-----HHGKDLNLLKLPFGTYTDFDLFLYDEE 1158
 + + + + ++ Q H + + L D F+ E
 Sbjct: 1138 TDGNYQTKETLKNLEMTQSDTGLVINQAQLAVVHRNQPSQLT----KMNQDFFISPNE 1193

Query: 1159 RANLISPKSVTVTISEKDSLKDVLKVNLLKKAALLVEFDKLLP-----KGATVQLVTKT 1213
 N K K+++ + L VN+ K + K P GA+V + T
 Sbjct: 1194 DGN-----KDFVAFKGLKNVYNDL-TVNVYAKD-----DHQKQTPWSSQAGASVSAIEST 1244

Query: 1214 NTVVDLPKATYSPTDYGNIPVGDYRLNVTLPSTYSLNLDLLVSVKEDQVNL--KL 1271
 A Y T G + GDY+ VT + E+ +SV + + +T +
 Sbjct: 1245 -----AWYGITARGSKVMPGDYQYVVTYRDEHGK-EHQKQYTSVNDKKPMITQGRF 1295

Query: 1272 TLINK----APLINALAEQTDIITQPVFYNAGTHLKNNYLANLEKAQTLIKNRVEQTSID 1327
 IN P + + I+ + VFY A KN + + + + I T D
 Sbjct: 1296 DTINGVDHFTPDKTALDSSGIVREEVFYLA---KKNRKFVDVTEGKDI-----TVSD 1346

-2303-

Query: 1328 NATAALRESQALNGKETDTSLLAKAILAETEIKGNYQFVNASPL----SQSTYIN---- 1379
 N + + + + D L+ + GN F L +N
 Sbjct: 1347 NKVIYIPKNEPDGSYTIISKRDGVTLSDYIYLVEDRAGNVSFATLRDLKAVGKDKAVVNFGLD 1406

Query: 1380 -QVQLAKNLLQKPNVTQSEVDKALENLDIAKNQLNGHETDYS--GLHHMIKANVLKQTS 1436
 V K ++ + + K +ENL+ N N Y + + N K S
 Sbjct: 1407 LPVPEDKQIVNFTYLVRDADGKPIENLEYNNNSGNSLILPYGKYTVELLTYDTNAAKLES 1466

Query: 1437 SKYQNASQFAKENYNNLIKKAELLSSNR-----QATQAQVEELINQIKATEQEL- 1485
 K + + A N+ + K +L +++ + ++ ++ +Q+ EQ L
 Sbjct: 1467 DKIVSFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLIPLEQSLY 1526

Query: 1486 -----DGRDRVSSAENYSQSLNDNDLSLNTTTPINPPNQPALIFKKGMTKES---- 1531
 +G V + + N +NT P N ++ + K G +S
 Sbjct: 1527 VPKAYGKTVQEGTYEVVVSLEPKGYRIEGNTKVNLTLP-NEVHELSSLRLVKVGDSSTGDH 1585

Query: 1532 -----EVAQKRVLGVTSTQTDNQKVKTNKLPKTGESTPKITYTILFSLSLMLGLATI 1582
 +Q T LP TGE K+ + + L +LGL +
 Sbjct: 1586 KVMSKNNQALTASATPTKSTTSATAKALPSTGE---KMGLKLRIVGLVLLGLTCV 1638

SEQ ID 8964 (GBS92) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 2; MW 48kDa).

GBS92-His was purified as shown in Figure 199, lane 9 .

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2038

A DNA sequence (GBSx2149) was identified in *S.agalactiae* <SEQ ID 6299> which encodes the amino acid sequence <SEQ ID 6300>. This protein is predicted to be AzlC family protein. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.80	Transmembrane	212 - 228 (196 - 230)
INTEGRAL	Likelihood = -7.27	Transmembrane	167 - 183 (159 - 185)
INTEGRAL	Likelihood = -5.68	Transmembrane	189 - 205 (188 - 210)
INTEGRAL	Likelihood = -2.28	Transmembrane	17 - 33 (13 - 34)
INTEGRAL	Likelihood = -1.06	Transmembrane	135 - 151 (135 - 151)
INTEGRAL	Likelihood = -1.01	Transmembrane	61 - 77 (60 - 77)

----- Final Results -----
 bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10235> which encodes amino acid sequence <SEQ ID 10236> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF10212 GB:AE001921 AzlC family protein [Deinococcus radiodurans]
 Identities = 72/224 (32%), Positives = 117/224 (52%), Gaps = 8/224 (3%)

Query: 6 FKEGVKDALPTALGYISIGLAFGIVASASDLAIEVGLMSALVYGGSAQFAMCALLLAKA 65
 F +G + +P LG + LA+ + A A+ LS + LMS + G++QFA L A A
 Sbjct: 7 FWQGFRLVPLWLGTVPFALAYAVTARAAGLSVGDTCLMSLTTFAGASQFAAAGLFGAHA 66

Query: 66 DLMTITMTVFLVNLRLNMLMSLHATTIFKSAHLMNQLAIGTLITDESYGV-LLGEALHHKV 124
 ++I +T FL+N R++L L + L ++ +TDE+YGV ++ A

-2304-

Sbjct: 67 GGLSIVLTTFLLNARHLLYGLSLARELRIT-LPQRVVAQAQFLTDEAYGVAVVSGARLPGG 125

Query: 125 VSPSWMHGNNVMSYLTWVISTIIGTLTGSTIPNPEMFGLDFALVAMFIGLFVFLFGMLS 184
 ++ +++ G + YL+W +ST++G L GS +P PE G+ F+GL V ++

5 Sbjct: 126 LTFAPLLGAELSLYLSWNVSTLLGALAGSVLPPEQLGVGVVFPFLAFLGLLV----PLVV 181

Query: 185 DGKRLVVYVLASVGLSYFLLATFLSGALSVLLATVVGCSVGVL 228
 D RL + V + GL + L+ L G L +LLA V G +G L

10 Sbjct: 182 D--RLSLLVALAAGLGGWALSRLVPGGLVILLAGVGGALLGAAL 223

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2039

15 A DNA sequence (GBSx2150) was identified in *S.agalactiae* <SEQ ID 6301> which encodes the amino acid sequence <SEQ ID 6302>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3794(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2040

30 A DNA sequence (GBSx2151) was identified in *S.agalactiae* <SEQ ID 6303> which encodes the amino acid sequence <SEQ ID 6304>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5087(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 10233> which encodes amino acid sequence <SEQ ID 10234> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04157 GB:AP001508 homosystein methyl transferase [Bacillus halodurans]
 Identities = 397/751 (52%), Positives = 519/751 (68%), Gaps = 14/751 (1%)

45 Query: 10 SNLGYPRIGEOREWKQAIEAFWAGNLEQKDLEKQLKQLRINHLKKQKEAGIDLIPVGDFS 69
 SNLGYPR+GE REWK+A+E+FWA + ++ L +K+LR+NHL+ Q+E +DLIPVGDF+
 Sbjct: 4 SNLGYPRIGENREWKKALESFWANDTTEEQLLATMKELRLNHLRVQQEQEVDLIPVGDF 63

50 Query: 70 CYDHVLDLSFQFNVIPKRFDEY--ERNLDLYFAIARGDKDNVASSMKKWFNTNYHYIVPE 127
 YDHVLD++ F +IPKRF + L YFA+ARG K+ A M KW+NTNYHYIVPE

-2305-

Sbjct: 64 LYDHVLDMAVMFGIIPKRFLQQGDTPTLSTYFAMARGSKNAQACEMTKWYNTNYHYIVPE 123

Query: 128 WEVETKPHLQNNYLLDLYLEAREVVGDKAKFVITGPITYVSLSSGIVD--FEATVQRLLP 185
+ P L N L+ YLEA+ +G KPVI GP ++V L+ G + + T+Q LLP

5 Sbjct: 124 LH-DAAPRLTKNAPLEAYLEAKNELGIDGKPVILGPYSFVKLAGYBEDKLQETIQSLLP 182

Query: 186 LYKQVFQDLIDAGATYIQIDEPIFVTDEGELLVDIAKSVYDFFAREVPQAHFIFQTYFES 245
LY QV Q+L+DAGA IQ+DEP VT + + +Y+ + A QTYF++

10 Sbjct: 183 LYIQVIQELVDAGARSIQVDEPSLVTSISAREMALVTRIYEQINEAIADAPLFLQTYFDA 242

Query: 246 AVCLDKLSKLPVTGFGLDIFIHGRAENLA AVKQ-GLFREKELFAGIVNGRNIWAVNLEETL 304
+++ LPV G GLDF+HG A+NL A++ G +K L ACI++GRNIW NL E

Sbjct: 243 VTFYEEVVS L P V K G I G L D F V H G A K N L E A L R T F G F P E D K V L A A G I I D G R N I W I S N L R E R H 302

15 Query: 305 ALLEEIGPFVK--RLTLQPSSSLHVPVTTKYETHLDPVLKNGLSFADEKLKELELLASA 362
L+ ++ V RL LQPS SLLHVPVTTK E LDP L L+FA+EKL EL L

Sbjct: 303 ELVHQLEQHVAKDRLLVLPQSCSLHVPVTTKREKLDPTLLGVLAFA NEKLT E L H T L K Q L 362

Query: 363 FDGNKTKGYHEALS R----FSALQAADFRIHVALESL-AEVKLESPYKLRQALQAEKLQL 417
GN+ + EAL +AL+ + +R A S E K + R+ LQ EK QL

20 Sbjct: 363 AAGNEAE-VKEALEANDDALAALEKSGWRSGAATSHNLENKKRPQSFNERRPLQEEKWQL 421

Query: 418 PILPTTTIGSFQSPQSPKIRKRLAWKRGNLSDSDYKDFIKTEIRRWIAIQEDLDLDVLVHG 477
P+LPTTTIGSFQ+ ++R+ R W++G LS +Y+ +K+ I +WI IQE+L LDVLVHG

25 Sbjct: 422 PLLPTTTIGSFQTKDVRRTSLWRKGELSTVEYERTMKSIEKWINIQEELGLDVLVHG 481

Query: 478 EFERVDMVEFFGQKLAGFTTTKLGWVQSYGSRVKKPPIIYGDVKHIQPLSLEETVYAQSL 537
EFER DMVEFFG+KL GF T GWVQSYGSR VKPPIIYG+V +P+++ ETVYAQSL

30 Sbjct: 482 EFERNDMVEFFGKLDGFAFTANGWVQSYGSRVKKPPIIYGNVSFTEPMTVAETVYAQSL 541

Query: 538 TKKPKVGMILTGPITITNWSFERDDISRSDLENQIALAIKDEIQLLEQSGIAIIQVDEAAL 597
T KPVKGMILTGP+TI NWSF RDD+ + + +QIA A+ E+ LE++GI +IQ+DE A+

Sbjct: 542 TDKPKVGMILTGPVTILNWSFVRDDLPLTVIAHQIAEALTHEVTALEEAGIEMIQIDEPAI 601

35 Query: 598 REGLPLRQKQQAAYLDDAVAAPKIAATSSVKDETQIHTHMCYSKFDEIIDSIRALDADVIS 657
REGLPL+ + QQ YLD AV+AF+ + + VK TQIHTHMCYS+F E+I++I LDADVIS

Sbjct: 602 REGLPLKAEDQQEYLDWAVSAFRASCAHVKATTQIHTHMCYSEFHEMIEAIDDLADVIS 661

40 Query: 658 IETSRSHGDIIESFETAVYPLGIGLVYDIHSPRIPTKEEIIVNIQRLKLSKEQFWVN 717
IETSRSHG++I +FE Y GIGLVYDIHSPR+P++EE++ I+R+L L FWN

Sbjct: 662 IETSRSHGEMISAFEKTTYEKIGIGLVYDIHSPRPSEEMLNVIRRALTVLPASLFWVN 721

Query: 718 PDCGLKTRREAETIAALEVLVSATKEVRQQL 748
PDCGLKTR E ET+AAL+ +V+A + R++L

45 Sbjct: 722 PDCGLKTRAEKETVAALKNMVAAARAAREEL 752

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50 Example 2041

A DNA sequence (GBSx2152) was identified in *S.agalactiae* <SEQ ID 6305> which encodes the amino acid sequence <SEQ ID 6306>. This protein is predicted to be methH. Analysis of this protein sequence reveals the following:

55 Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.0753 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2306-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05348 GB:AP001512 unknown conserved protein [Bacillus halodurans]
Identities = 301/610 (49%), Positives = 437/610 (71%), Gaps = 9/610 (1%)

5 Query: 1 MSKFLEKLTDLVADGAMGTLTYGLDTCHESTYNVTHPEKVLAIHQAYIEAGADVIQT 60
M+ +E LKT+ILV DGAMGTLTY G+D C E NVT PEK++A H AY+EAGADVIQT
Sbjct: 1 MTNLVEALKTNILVGDGAMGTLTYEQGIDRCFEELNVTDPKIVAAHVAYVEAGADVIQT 60

10 Query: 61 NTYGAQRHRLKNYGLEQVVSINQAAVNIAHQATLGKETFILGTGGRSQRQCDLTLDN 120
NTY A R +L Y L+DQV+ IN+AAV +A +A +ETF+LGT+GG RS + ++ +
Sbjct: 61 NTYANRMKLAKYQLDDQVLEINRAAVRLARKAAK-QETFLGTIGGIRSVQFEEVEIQE 119

15 Query: 121 IVEETLEQVEALLATGQDGLLFTYYDIEETTVLKIVREMTDLPIITNISLHEAGVTS 180
+ + LEQ++AL++ G +DGLL ET+YD+EE + + R +TDLP+I ++S+ E GV
Sbjct: 120 VQDVFLEQMKALVSEG-VDGLLLETFYDLEEKAVSLARSLTDLFVIAHLSTAEIGVLQ 178

20 Query: 181 NGKPIVEALSQVLMGADVIGLNCHLGPYHMIQSLKQVPLFAQSYLSVYPNASQLSLDGE 240
GK + EA ++L LGAD++G+NC +GPY M++SL+ V L ++Y S YPNAS D
Sbjct: 179 GKKLLEEAFALEGLGADLVGINCRMGYPQMLRSLETVQLLDRAVYSAYPNASLP--DYP 236

25 Query: 241 NSQYQFSQNSEYFGKSAELLVAEGVRLIGGCCGTTDPHIRAVKRSIRGLKPIERKVVTP 300
+ + + N EYF + + V +GVRL+GCCCGTTP+H+RA + ++GLKP+ K V
Sbjct: 237 DGRLYYHSNPEYFYEMGKRFFVQCGVRLIGGCCGTTPEHVRFAKVVVKGLKPVVSKPVR-- 294

30 Query: 301 IPVKDFVRRIRRT---DTLVKVKKEVTIIAELDPPKHLDIVQFKAIRAIDQKGIAAIT 357
+ +K+ + + + L +KVKK+ +II ELDPPK+L I +F + A+ G+ A+T
Sbjct: 295 LEIKETLSSTGQKTAREPLAEKVKKQPSIIVELDPPKNLAIDRFVEGAAALKNAGVDAVT 354

35 Query: 358 LADNSLSNTRICNLISIASLLKDEISTPFLHIACRDHNLIQLSRLGMEILLGFNHILAI 417
+ADNSL++ R+ NL++ ++++ ++ L+H+ CRD NLIGLQS L+G+ LG +LAI
Sbjct: 355 MADNSLASPRVDNLALGAIITQQVGARPLVHVTCRDRNLIGLQSHLMGLHALGMTDLAI 414

40 Query: 418 TGDPTKLGDFPGATSVYDVTSFKLKSLIKQLNQLSYSGASLRRPTDFTVAAAFNPVNKN 477
TGDPTK+GDFPGATSVYDVTSL+L+SLIKQLN+G+S+SG L + +F+V AAFNPVN++
Sbjct: 415 TGDPTKVGDFPGATSVYDVTSLQLISLIKQLNEGISFSGKELGQKANFSVGAAFPNPNVRH 474

45 Query: 478 LTRTVKLEKKVASGADYFMTQPIFDHSLVKELADLTKTVEQFFFIGIMPITSYNNAVFL 537
L R V+ +EKK+ +GADYFMTQPI++ ++++ + TK +E+P +IGIMP+ + NA FL
Sbjct: 475 LERAVQRMKKIEAGADYFMTQPIYNEKQIEDIYEATKHIEKPIYIGIMPLINGRNAEFL 534

50 Query: 538 HNEVPGIKLSSESFLSALEKVKDDKEACLTALNESKSLIDEALNYFNIGIYLITPFLRYDL 597
HNEVPGIKL++ + + +D++ L +KSL+D A +YFNIGIYLITPFLRY +
Sbjct: 535 HNEVPGIKLTDQIRERMARAGEDRQKGEREGLAIAKSLLDVATHYFNIGIYLITPFLRYGM 594

55 Query: 598 TLELIDYIQK 607
T++L Y+++
Sbjct: 595 TVDLTHYVKE 604

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2042

A DNA sequence (GBSx2153) was identified in *S.agalactiae* <SEQ ID 6307> which encodes the amino acid sequence <SEQ ID 6308>. Analysis of this protein sequence reveals the following:

55 Possible site: 53
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -9.55 Transmembrane 127 - 143 (121 - 147)
INTEGRAL Likelihood = -1.44 Transmembrane 157 - 173 (155 - 175)

60 ----- Final Results -----
bacterial membrane --- Certainty=0.4821(Affirmative) < succ>

-2307-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10231> which encodes amino acid sequence <SEQ ID 10232> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC01354 GB:AL390975 putative integral membrane protein
  [Streptomyces coelicolor A3(2)]
Identities = 38/98 (38%), Positives = 59/98 (59%)

Query: 113 RIADDVARFGGSWTFIIVFVSIMAIWMLVNIMKPFQIQFDPYPFILLNLALSTIAAIQAP 172
      R+++ VARF G+  FI+  ++ +W++ N+  P G++FD YPFI L L LS A+  AP
Sbjct: 47  RLSERVARFLGTGRFIVWMTVVIIWVWVNVSAFSGLRFDYPPFIFLTLMLSLQASYAAP 106

Query: 173 LIMMSQNRAADYDRLQARNDNFVNKTSELEIRLLHEKI 210
      LI+++QNR D DR+  D  N+ S +  L +I
Sbjct: 107 LILLAQNRQDDRDVNLQDRKQNSIADTEYLTREI 144
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8965> and protein <SEQ ID 8966> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1  Crend: 7
McG: Discrim Score: -3.84
GvH: Signal Score (-7.5): -5.05
      Possible site: 53
>>> Seems to have no N-terminal signal sequence
ALOM program count: 2 value: -9.55 threshold: 0.0
INTEGRAL Likelihood = -9.55 Transmembrane 127 - 143 ( 121 - 147)
INTEGRAL Likelihood = -1.44 Transmembrane 157 - 173 ( 155 - 175)
PERIPHERAL Likelihood = 5.46 27
modified ALOM score: 2.41

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01598(637 - 930 of 1341)
GP|9714438|emb|CAC01354.1|AL390975(47 - 144 of 198) putative integral membrane protein
{Streptomyces coelicolor A3(2)}
%Match = 8.2
%Identity = 38.8 %Similarity = 61.2
Matches = 38 Mismatches = 38 Conservative Subs = 22

600      630      660      690      720      750      780      810
MKEBEKPFNVVEERLNKQATIGQRIADDVARFGGSWTFIIVFVSIMAIWMLVNIMKPFQIQFDPYPFILLNLALSTIAAIQ
      |::: |||| | : ||:  :: :||: | : | |::|| ||||:| | | | :
RLDQPRPPRRLLPEWDPESFGRLSERVARFLGTGRFIVWMTVVIIWVWVNVSAFSGLRFDYPPFIFLTLMLSLQASYA
      40      50      60      70      80      90      100

840      870      900      930      960      990      1020      1050
APLIMMSQNRAADYDRLQARNDNFVNKTSELEIRLLHEKIDHMQDQFELLEIQKLQTEMLVSLGNQLAQLKQLK*SF
|||:::| | | | : | : | : |
APLILLAQNRRQDDRDVNLQDRKQNSIADTEYLTREIAALRIGLGEVATRDWIRSELQDLVRDLLEERQNGHHFDRGV
```

-2308-

120 130 140 150 160 170 180

SEQ ID 8966 (GBS393) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 3; MW 30.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 177 (lane 4; MW 56kDa) and in Figure 83 (lane 6; MW 56kDa).

GBS393-GST was purified as shown in Figure 217, lane 5.

Example 2043

A DNA sequence (GBSx2154) was identified in *S.agalactiae* <SEQ ID 6309> which encodes the amino acid sequence <SEQ ID 6310>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.29 Transmembrane 274 - 290 (271 - 291)
----- Final Results -----
 bacterial membrane --- Certainty=0.2317(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35508 GB:AE001721 glycerol dehydrogenase [Thermotoga maritima]
Identities = 94/307 (30%), Positives = 157/307 (50%), Gaps = 21/307 (6%)

Query: 63 VYGTSTQSNIDKLIVANPQVQAADAILGFGGKALDTAKMVAKELGKNSFTIPTICSNC 122
 ++G + + I++L + + D ++G GGGK LDTAK VA +L K +PTI S +
Sbjct: 62 IFGGECSDEBIERLSGLVE-BETDVVVGIGGGKTLDTAKAVAYKLKKPVVIVPTIASTDA 120

Query: 123 AGTAIAVVYNDHDSFLRYGY-PESPLHIFINTRIIAQAPSKYFWAGIGDGISKAPEVERA 181
 +A++V+Y + F RY + P +P + ++T I+A+AP+++ AG+GD ++ E E
Sbjct: 121 PCSALSVIYTPNGEFKRYLFLPRNPVVLVDTEIVAKAPARFLVAGMGDALATWFEAESC 180

Query: 182 TLEAKTNKLPHT-AVLGQAVALLSSKEAFYQFGEQGLKDVEANLASRAVEEI--ALDILIS 238
 + N ++ A+A E ++G + VE + A+E+I A +L
Sbjct: 181 KQKYAPNMTGRLGSMYALARLCYETLLEYGVLAKRSVEEKSVPAPLEKIVEANTLLSG 240

Query: 239 TGYASNLVNQPDFYNSCHAHAFYYGTTAIQRQGEFLHGVVVAFGVLV-LHAYFNELEBEL 297
 G+ S AHA + G T ++ ++LHG VA GVL L + +
Sbjct: 241 LGFESC-----GLAAAHAIHNGLTVLENTHKYLGHEKVAIGVLASLFLTDKPRKMI 291

Query: 298 EKVARFNKSLGLPTTLADVSL---SEKDIPKIVEIAMTNE---YKNTPFDPKMFQAAIL 351
 E+V F + +GLPTTLA++ L S++D+ K+ E A NE + P K A+
Sbjct: 292 EEVYSFCEEVGLPTTLAEIGLDGVSDEDLMKVAEKACDKNETIHNEPQPVTSKDVFFALK 351

Query: 352 AADAFGQ 358
 AAD +G+
Sbjct: 352 AADRYGR 358

There is also homology to SEQ ID 3078.

SEQ ID 6310 (GBS123) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 7; MW 43.3kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2309-

Example 2044

A DNA sequence (GBSx2155) was identified in *S.agalactiae* <SEQ ID 6311> which encodes the amino acid sequence <SEQ ID 6312>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0974(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6313> which encodes the amino acid sequence <SEQ ID 6314>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2368(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 92/167 (55%), Positives = 121/167 (72%)

Query: 1 MKIAIIGYSGSGKSTLARKLGNYYNCNVLHLD SIHFAPNWEERKYDDMIDDVSNMLEKRT 60
 +KIAIIG+SGSGKSTLAR LG +Y+C V HLD +HF+ NW+ER DMI D+S L K+
 Sbjct: 1 LKIAIIGHSGSGKSTLARFLGQHYHCEVFHLDQLHFSSNWQERSDHDMIADLSTCLLKQD 60

Query: 61 WIEGNYKKLLYQERLADADEIIFDFNRFNCLWRAFKRYCKFRGKTRPDMANGCPEKLD 120
 IIEGNY LY+ER+++AD II+ +F+RF+C++RAFKRY +RGKTRPDMA+ C EK D
 Sbjct: 61 LIIEGNYANCLYEERMSEADYIIYVNFSRFHCYVYAFKRYLNYRGKTRPDMADNCQEKFD 120

Query: 121 FEFISWILKDGSRDKQKSNYKQVVEDYPQKIKILKHQRDLDDQYLKEL 167
 F+ WIL DGRS Q Y+ VV+ Y K +L +Q+ L Y+ +
 Sbjct: 121 VAFVKWILLDGRSRNQLKKYQSVVQKYSHKTIIVLTNQQLSHYMNTI 167

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2045

A DNA sequence (GBSx2156) was identified in *S.agalactiae* <SEQ ID 6315> which encodes the amino acid sequence <SEQ ID 6316>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3874(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA41941 GB:X59250 initiation factor IF-1 [Lactococcus lactis]
 Identities = 62/72 (86%), Positives = 70/72 (97%)

Query: 1 MAKEDVIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKKNYIRILVGDRVTVMSPY 60
 MAK+DVIE++GKVV+TMPNAMFTVELENGHQ+LAT+SGKIRKKNYIRIL GD+V VE+SPY
 Sbjct: 1 MAKDDVIEVDGKVVDTMPNAMFTVELENGHQVLATISGKIRKKNYIRILPGDKVQVBLSPY 60

-2310-

Query: 61 DLTRGRITYRFK 72
DLTRGRITYRFK
Sbjct: 61 DLTRGRITYRFK 72

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6317> which encodes the amino acid sequence <SEQ ID 6318>. Analysis of this protein sequence reveals the following:

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3253(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 67/67 (100%), Positives = 67/67 (100%)

Query: 6 VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKKNYIRILVGDRVTVEMSPYDLTRG 65
VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKKNYIRILVGDRVTVEMSPYDLTRG
Sbjct: 1 VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKKNYIRILVGDRVTVEMSPYDLTRG 60

Query: 66 RITYRFK 72
RITYRFK
Sbjct: 61 RITYRFK 67

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2046

A DNA sequence (GBSx2157) was identified in *S.agalactiae* <SEQ ID 6319> which encodes the amino acid sequence <SEQ ID 6320>. This protein is predicted to be adenylate kinase (adk). Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA41940 GB:X59250 adenylate kinase [Lactococcus lactis]
Identities = 146/214 (68%), Positives = 170/214 (79%), Gaps = 6/214 (2%)

Query: 1 MNLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRILAKSYIDKGELVP 60
MNLLIMGLPGAGKGTQA IV+ +GV HISTGDMFRAAM N+TEMG+LAKS+IDKGELVP
Sbjct: 1 MNLLIMGLPGAGKGTQAEFIVKNYGVNHISTGDMFRAAMKNETEMGKLAKSFIDKGELVP 60

Query: 61 DEVTNGIVKERLAEDDIAEKGFLLDGYPRITIEQAHALDATLEELGLRLDGVINIKVDPSC 120
DEVTNGIVKERLA+DDI GFLLDGYPRIT+QAHALD LEELG++LD V+NI V+P+
Sbjct: 61 DEVTNGIVKERLAQDDIKASGFLLDGYPRITIDQAHALDTMLEELGIKLDVAVNIVNPN 120

Query: 121 LIERLSGRIINRKTGETFHKVFNPVV-----DYKEEDYYQREDDKPVKRRLDVNI 174
L++RLSGR I R G T+HK+FNP D YQR DD PETVK RLDVNI +

Sbjct: 121 LVDRLSGRYICRNCGATYHKIFNPPTKVEGTCDVCGSHDLYQRADDVPETVKNRLDVNIKE 180

Query: 175 GEPILEHYRKLGLVTDIEGNQEITEVFADVEKAL 208
PI+EHY +LGLV +IEG QEI++V D++K L

-2311-

Subject: 181 SAPIIEHYTELGLVKNIEGEQEISOVTDDIKKVL 214

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6321> which encodes the amino acid sequence <SEQ ID 6322>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 17
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
                                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10                                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 208/212 (98%), Positives = 212/212 (99%)

15

Query: 1 MNLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVP 60
MNLLIMGLPGAGKGTQAAKIVEEFG+AHISTGDMFRAAMANQTEMGR LAKSYIDKGELVP
Sbjct: 1 MNLLIMGLPGAGKGTQAAKIVEEFGIAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVP 60

20

Query: 61 DEVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC 120
DEVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC
Sbjct: 61 DEVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC 120

25

Query: 121 LIERLSGRIINRK TGETFHKVFNPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILE 180
L+ERLSGRIINRK TGETFHKVFNPVDYKEEDYYQREDDKPETVKRRLDVN+AQGEPILE
Sbjct: 121 LVERLSGRIINRK TGETFHKVFNPVDYKEEDYYQREDDKPETVKRRLDVNMAQGEPILE 180

30

Query: 181 HYRKLGLVTDIEGNQ EITVFADVEKALLELK 212
HYRKLGLVTDIEGNQ EIT+VFADVEKALLELK
Sbjct: 181 HYRKLGLVTDIEGNQ EIT+VFADVEKALLELK 212

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8967> and protein <SEQ ID 8968> were also identified. Analysis of this
35 protein sequence reveals the following:

```

Lipop: Possible site: -1      Crend: 0
McG: Discrim Score:      -1.04
GvH: Signal Score (-7.5): -1.08
      Possible site: 17
40 >>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value: 6.79 threshold: 0.0
      PERIPHERAL Likelihood = 6.79      106
      modified ALOM score: -1.86

45 *** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

over 213aa

55 EGAD|8612| adenylate kinase Insert characterized Lactococcus lactis
SP|P27143|KAD_LACLA ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP. TRANSPHOSPHORYLASE). Edit
characterized
GP|44074|emb|CAA41940.1||X59250 adenylate kinase Insert characterized
PIR|S17987|S17987 adenylate kinase (EC 2.7.4.3) - subsp. lactis Insert characterized
60 PIR|B44812|B44812 adenylate kinase (EC 2.7.4.3) - Insert characterized

[illegible]

The GBS114-His fusion product was purified (Figure 108A; see also Figure 200, lane 8) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 108B), FACS (Figure 108C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

A DNA sequence (GBSx2158) was identified in *S.agalactiae* <SEQ ID 6323> which encodes the amino acid sequence <SEQ ID 6324>. This protein is predicted to be preprotein translocase secY subunit (secY). Analysis of this protein sequence reveals the following:

50	INTEGRAL	Likelihood	= -14.01	Transmembrane	217 - 233 (209 - 240)
	INTEGRAL	Likelihood	= -8.65	Transmembrane	314 - 330 (307 - 334)
	INTEGRAL	Likelihood	= -6.16	Transmembrane	369 - 385 (363 - 392)
	INTEGRAL	Likelihood	= -5.36	Transmembrane	19 - 35 (17 - 40)
55	INTEGRAL	Likelihood	= -3.93	Transmembrane	180 - 196 (179 - 199)
	INTEGRAL	Likelihood	= -3.03	Transmembrane	395 - 411 (392 - 412)
	INTEGRAL	Likelihood	= -2.55	Transmembrane	151 - 167 (151 - 168)
	INTEGRAL	Likelihood	= -2.02	Transmembrane	117 - 133 (117 - 133)
	INTEGRAL	Likelihood	= -0.64	Transmembrane	270 - 286 (269 - 286)

----- Final Results -----

-2313-

bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9467> which encodes amino acid sequence <SEQ ID 9468> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA41939 GB:X59250 SecY protein [Lactococcus lactis]
 Identities = 292/433 (67%), Positives = 361/433 (82%), Gaps = 2/433 (0%)

10 Query: 1 MFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLQMGELPFLNMLNLV 60
 MF K L++A KVK VR +ILFTIFIL VFR+G HIT PG+NV++L+Q+ +LPFL+M+NLV
 Sbjct: 1 MFFKTLKEAFKVKDVRARILFTIFILFVFRILGAHITAPGVNVQNQQVADLPFLSMMNLV 60

15 Query: 61 SGNAMRNFSVFSMGVSPYITASIVVQLQMDILPKFVEWGKQGEVGRRLNQATRYISLF 120
 SGNAM+N+S+F+MGVSPYITASI+VQLQMDILPKFVEW KQGE+GRRRLNQATRYI+L
 Sbjct: 61 SGNAMQNYSLFAMGVSPYITASIVVQLQMDILPKFVEWSKQGEIGRRRLNQATRYIITLV 120

20 Query: 121 LAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIGAILTTGSMVVTWLGEQITDKGFGNGV 180
 LA QSIGITAGF +SS+ +V+ PN Q+YL+IG +LTIGSMVVTW+GEQI +KGFG+GV
 Sbjct: 121 LAMAQSIGITAGFQAMSSNLIVQNPWQSYLMIGVLLTTGSMVVTWMGEQINEKGFSGV 180

25 Query: 181 SMIIFAGIISIPSAITTIYEDFFVNVRSSAITNSYIFVGILIVAVLAI VFFTTFIQQAE 240
 S+IIFAGI+S IPSAI ++Y++ F+NVR S I S+IFV LI++ + I++ TTF+QQAE
 Sbjct: 181 SVIIFAGIVSGIPSAIKSVYDEKFLNVRPSEIPMSWIFVIGLILSAIVIIYVTTTFVQQAE 240

30 Query: 241 YKIPIQYTKLVQGAPTSSYLPLKVNPAAGVIPVIFASSITTIPTIIPFFQ--NGKEIPWL 298
 K+PIQYTKL QGAPTSSYLPL+VNPAGVIPVIFA SITT P+TI+ F Q G + WL
 Sbjct: 241 RKVPIQYTKLTQGAPTSSYLPLRVNPAAGVIPVIFAGSITTAPATILQFLQRSQGSNVGWL 300

35 Query: 299 TKLQELLNYQTPVGMIIYAILIILFSFFYTFVQVNPEKTAENLQKNSSYIPSIRPGRETE 358
 + LQ L+Y T GM+ YA+LI+LF+FFY+FVQVNPEK AENLQK SYIPS+REG+ TE
 Sbjct: 301 STLQNALSYTTWTGMLFYALLIVLFTFFYSFVQVNPEKMAENLQKGSYIPSVRPGKGT 360

40 Query: 359 EYMSSLLKKLATIGSVFLAFISLLPIIAQQALHLSSSIALGGTSLLLILIATGIEGMKQLE 418
 +Y+S LL +LAT+GS+FL IS++PI AQ L +ALGGTSLLLILI I+ +KQLE
 Sbjct: 361 KYVSRLLMRLATVGSFLGLISTIPIAAQNVWGLPKIVALGGTSLLLILIQVAIQAVKQLE 420

45 Query: 419 GYLLKRRYVGFMN 431
 GYLLKR+Y GFM+
 Sbjct: 421 GYLLKRRYAGFMD 433

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3987> which encodes the amino acid sequence <SEQ ID 3988>. Analysis of this protein sequence reveals the following:

45 Possible site: 55
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -14.70	Transmembrane	233 - 249 (226 - 255)
INTEGRAL	Likelihood = -8.12	Transmembrane	330 - 346 (323 - 350)
INTEGRAL	Likelihood = -6.10	Transmembrane	384 - 400 (378 - 403)
50 INTEGRAL	Likelihood = -5.20	Transmembrane	35 - 51 (33 - 56)
INTEGRAL	Likelihood = -4.09	Transmembrane	199 - 215 (195 - 215)
INTEGRAL	Likelihood = -3.56	Transmembrane	167 - 183 (165 - 184)
INTEGRAL	Likelihood = -1.65	Transmembrane	411 - 427 (411 - 428)
INTEGRAL	Likelihood = -1.49	Transmembrane	133 - 149 (133 - 149)
55 INTEGRAL	Likelihood = -0.64	Transmembrane	286 - 302 (285 - 302)

----- Final Results -----

bacterial membrane --- Certainty=0.6880(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60

An alignment of the GAS and GBS proteins is shown below.

-2314-

Identities = 377/434 (86%), Positives = 417/434 (95%)

Query: 1 MFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLEQMGEPLPFLNMLNLV 60
 MFLK+L+DALK+K VRNKI FTIFI+LVFRIGTHITVPG+N KSLEQ+ ELPPFLNMLNLV
 5 Sbjet: 17 MFLKILKDALKIKITVRNKIFFTIFILLVFRIGTHITVPGVNAKSLEQLSELPPFLNMLNLV 76

Query: 61 SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRLNQATRYISLF 120
 SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRLNQATRYISL
 10 Sbjet: 77 SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRLNQATRYISLV 136

Query: 121 LAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIGAILTTGSMVVTWLGEQITDKGFGNGV 180
 LAF QSIGITAGFNTLS+VALVKTP+++TYLLIGA+LTTGS++VTWLGEQITDKGFGNGV
 15 Sbjet: 137 LAFVQSIGITAGFNTLSNVALVKTPDIKTYLLIGALLTTGSMVVTWLGEQITDKGFGNGV 196

Query: 181 SMIIIFAGIISSIPSAITTIYEDFFVNVRRSAITNSYIFVGILIVAVLAIVFFTTFIQQAE 240
 SMIIIFAGIISSIPSAI TI ED+FVNV++S + +SY+ VGILI+AVLAIVFFTT++QQAE
 20 Sbjet: 197 SMIIIFAGIISSIPSAIATIREDFVNVKASDLHSSYLIVGILIIAVLAIVFFTYVQQAE 256

Query: 241 YKIPIQYTKLVQGAFTSSYLPLKVNPAAGVIPVIFASSITTIPSTIIPFFQNGKEIPWLTK 300
 YKIPIQYTKL+QGAFTSSYLPLKVNPAAGVIPVIFASSITTIPSTIIPF QNG+++PWL +
 25 Sbjet: 257 YKIPIQYTKLMQGAFTSSYLPLKVNPAAGVIPVIFASSITTIPSTIIPFFQNGRDLPLWLN 316

Query: 301 LQELLNYQTPVGMIIYAILIILFSFFYTFVQVNPEKTAENLQKNSSYIPSRPGRETEEY 360
 LQE+ NYQTPVGMIL+YA+LIILFSFFYTFVQVNPEKTAENLQKNSSYIPS+RPGRETE++
 30 Sbjet: 317 LQEIFNYQTPVGMIMIVYALLIILFSFFYTFVQVNPEKTAENLQKNSSYIPSVRPGRETEQF 376

Query: 361 MSSLLKKLATIGSVFLAFISLLPIIAQQALHLSSSIALGGTSLILLIATGIEGMKQLEGY 420
 MS+LLKKLAT+G++FLAFISL PI AQAL+LSSSIALGGTSLILLI+TGIEGMKQLEGY
 35 Sbjet: 377 MSALLKKLATVGAIFLAFISLAPIAAQQALNLSSSIALGGTSLILLISTGIEGMKQLEGY 436

Query: 421 LLKRRYVGFMMNTTE 434
 LLKR+YVGFMMNT E
 40 Sbjet: 437 LLKRRYVGFMMNTAE 450

35 A related GBS gene <SEQ ID 8969> and protein <SEQ ID 8970> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 6.16
 GvH: Signal Score (-7.5): -4.32
 40 Possible site: 35
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 9 value: -14.01 threshold: 0.0

INTEGRAL	Likelihood = -14.01	Transmembrane	217 - 233 (209 - 240)
INTEGRAL	Likelihood = -9.98	Transmembrane	311 - 327 (307 - 334)
INTEGRAL	Likelihood = -6.16	Transmembrane	369 - 385 (363 - 392)
INTEGRAL	Likelihood = -5.36	Transmembrane	19 - 35 (17 - 40)
INTEGRAL	Likelihood = -3.93	Transmembrane	180 - 196 (179 - 199)
INTEGRAL	Likelihood = -3.03	Transmembrane	395 - 411 (392 - 412)
INTEGRAL	Likelihood = -2.55	Transmembrane	151 - 167 (151 - 168)
INTEGRAL	Likelihood = -2.02	Transmembrane	117 - 133 (117 - 133)
INTEGRAL	Likelihood = -0.64	Transmembrane	270 - 286 (269 - 286)
PERIPHERAL	Likelihood = 0.95		69

modified ALOM score: 3.30

55 *** Reasoning Step: 3

----- Final Results -----

bacterial membrane	--- Certainty=0.6604 (Affirmative) < succ>
bacterial outside	--- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000 (Not Clear) < succ>

60

The protein has homology with the following sequences in the databases:

ORF01657 (301 - 1596 of 1902)
 EGAD|6545|6344 (1 - 434 of 439) preprotein translocase secy subunit {Lactococcus lactis}
 65 SP|P27148|SECY_LACLA PREPROTEIN TRANSLOCASE SECY SUBUNIT. GP|44073|emb|CAA41939.1|X59250

-2315-

SecY protein {Lactococcus lactis} PIR|S17985|S17985 preprotein translocase secY -
 Lactococcus lactis subsp. lactis
 %Match = 46.6
 %Identity = 67.0 %Similarity = 84.1
 Matches = 290 Mismatches = 68 Conservative Sub.s = 74

```

72      102      132      162      192      222      252      282
HQCKRICSECP*PIKCL*RWY*SNSSCS*RSWNRAC*KIRR*NSW*W*IN*EIVC*SS*IF*IC*SSYHC*RFWFNRSHLI

10      312      342      372      402      432      462      492      522
NER*LIMFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLQMGELPFLNMLNLVSGNAMRNFVSFMSG
||:| |:|:| ||| |:| ||||| |||:| ||| |:|:| |:|: |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
MFEKTLKEAFKVKDVRARILFTIFILFVFRLGAIHTAPGVNVQNQQVADLPFLSMMNLVSGNAMQNSYSLFAMG

15      10      20      30      40      50      60      70
552      582      612      642      672      702      732      762
VSPYITASIVQQLQMDILPKFVWEGKQGEVGRKLNQATRYISLFLAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIG
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
VSPYITASIVQQLQMDILPKFVWEGKQGEIGRRKLNQATRYITLVLAMAQSIGITAGFQAMSSLNIVQNPWQSYLMIG

20      90      100      110      120      130      140      150
792      822      852      882      912      942      972      1002
AILTTGSMVVTWLGEQITDKGFGNGVSMIIFAGIISIPSAITTIYEDFFVNVRRSSAITNSYIFVGILIVAVLAIVFFTT
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
VLLTTGSMVVTWMGEQINEKGFGSGVSVIIFAGIVSGIPSAIKSVYDEKFLNVRPSEIPMSWIFVIGLILSAIVIIYVTT

25      170      180      190      200      210      220      230
1032      1062      1092      1122      1152      1176      1206      1236
FIQQAERYKIPIQYTKLVQGAFTSSYLPLKVNPAQVIVPFIASSITITPSTIIPFFQ--NGKEIPWLTKLQELLNQTTPVG
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
FVQQAERKVPKIQYTKLVQGAFTSSYLPLRVNPAQVIVPFIAGSITTAPATILQFLQRSQSSNVGWLSTLQNALSYTTWTG

30      250      260      270      280      290      300      310
1266      1296      1326      1356      1386      1416      1446      1476
MIIYAILIILFSFFXTFXQVNPEKTAENLQKNSSYIPSIIRPGRETEEYMSLLKKLATIGSVFLAFISLLPIIAQQALHL
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
MLFYALLIVLFTFFYSFVQVNPEKMAENLQKQGSYIPSVRPGKGTEKYVSRLLMRLATVGSFLGLISIIPIAAQNVWGL

35      330      340      350      360      370      380      390
1506      1536      1566      1596      1626      1656      1686      1716
SSSIALGGTSLILILATGIEGMKQLEGYLLKRRYVGFMMTTE*NIG*LCQPSILFFNKSDMLCWIYLLKTK*GDYNESFNY
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
PKIVALGGTSLILIQVAIQAVKQLEGYLLKRRYAGFMDNPLETK

40      410      420      430

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2048

A DNA sequence (GBSx2159) was identified in *S. agalactiae* <SEQ ID 6325> which encodes the amino acid sequence <SEQ ID 6326>. This protein is predicted to be 50S ribosomal protein L15 (rplO). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5259(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB54021 GB:U96620 ribosomal protein L15 [Staphylococcus aureus]

-2316-

Identities = 116/146 (79%), Positives = 128/146 (87%)

Query: 1 MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60
 Sbjct: 1 MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60

Query: 61 MPKRGFSNINAKKEYALVNLDQLNVFEDGTEVTPVVLKEAGIVRAEKSGVKILGNGELTKK 120
 +PKRGF+NIN KEYA+VNLDQLN FEDGTEVTP +L E+G+V+ EKSG+KILGNG L KK
 Sbjct: 61 LPKRGFTNINRKEYAIVNLDQLNKFEDGTEVTPALLVESGVVKNEKSGIKILGNGSLDKK 120

Query: 121 LSVKAAKFSKSAEAAITAKGGSIEVI 146
 L+VKA KFS SA AI AKGG+ EVI
 Sbjct: 121 LTVKAHKFSASAAEAIDAKGGAHEVI 146

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6327> which encodes the amino acid sequence <SEQ ID 6328>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5329(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 135/146 (92%), Positives = 142/146 (96%)

Query: 1 MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60
 Sbjct: 1 MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60

Query: 61 MPKRGFSNINAKKEYALVNLDQLNVFEDGTEVTPVVLKEAGIVRAEKSGVKILGNGELTKK 120
 +PKRGF+NIN KEYALVNLDQLNVF+DGTEVTP +LK+AGIVRAEKSGVK+LGNGELTKK
 Sbjct: 61 IPKRGFTNINTKEYALVNLDQLNVFDDGTEVTPAILKDAGIVRAEKSGVKVLGNGELTKK 120

Query: 121 LSVKAAKFSKSAEAAITAKGGSIEVI 146
 L+VKAHKFSKSAEAAI AKGGSIIEVI
 Sbjct: 121 LTVKAHKFSKSAEAAITAKGGSIEVI 146

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2049

A DNA sequence (GBSx2160) was identified in *S.agalactiae* <SEQ ID 6329> which encodes the amino acid sequence <SEQ ID 6330>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1162(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB54020 GB:U96620 ribosomal protein L30 [Staphylococcus aureus]
 Identities = 40/58 (68%), Positives = 46/58 (78%)

Query: 1 MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMVNAISHLVTVEE 58
 MA+++ITLT+S IGR QRKTV ALGL K NSSVV EDN AIRG +N + HLTVEE

-2317-

Sbjct: 1 MAKLIKITLRSVIGRPETQRKTVEALGLKKTNSVVVEDNPAIRGQINKVKHLVTVEE 58

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6331> which encodes the amino acid sequence <SEQ ID 6332>. Analysis of this protein sequence reveals the following:

5 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 56/58 (96%), Positives = 57/58 (97%)

 Query: 1 MAQIKITLTKSPIGRKPEQRKTVALGLGKLNSSVVKEDNAAIRGMVNAISHLVTVEE 58
 MAQIKITLTKSPIGRKPEQRKTVALGLGKLNSSVVKEDNAAIRGMV AISHLVTVE+
 Sbjct: 1 MAQIKITLTKSPIGRKPEQRKTVALGLGKLNSSVVKEDNAAIRGMVTAISHLVTVED 58

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2050

A DNA sequence (GBSx2161) was identified in *S.agalactiae* <SEQ ID 6333> which encodes the amino acid sequence <SEQ ID 6334>. Analysis of this protein sequence reveals the following:

25 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.3226(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2051

40 A DNA sequence (GBSx2162) was identified in *S.agalactiae* <SEQ ID 6335> which encodes the amino acid sequence <SEQ ID 6336>. This protein is predicted to be 30S ribosomal protein S5 (rpsE). Analysis of this protein sequence reveals the following:

 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.3179(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAA22699 GB:M57621 ribosomal protein S5 [Bacillus
 stearothermophilus]

-2318-

Identities = 119/158 (75%), Positives = 139/158 (87%)

Query: 6 NAVELEERVVAINRVTKVVGGRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIRKA 65
 N +ELEERVVA+NRV KVVKGGRRLRF+ALVVVGD+NG VGFGTGKAQEVPEAIRKA+E
 5 Sbjct: 7 NKLELEERVVAVNRVAKVVGGRRLRFSALVVVGDKNHVGFGTGKAQEVPEAIRKAIED 66

Query: 66 AKKNMVEVPMVGTTPHEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVADITSK 125
 AKKN++EVP+VGTTPHEV FG +++LKPA EG GV AGG RAV+ELAG++DI SK
 10 Sbjct: 67 AKKNLIEVPIVGTTPHEVIGHFGAGEIILKPASEGTGVIAGGPARAVLELAGISDILSK 126

Query: 126 SLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDI 163
 S+GSNTPIN+VRAT +GLKQLKRAE+VA LRG +V +L
 Sbjct: 127 SIGSNTPINMVRATFDGLKQLKRAEDVAKLRGKTVEEL 164

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6337> which encodes the amino acid sequence <SEQ ID 6338>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3179(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 158/164 (96%), Positives = 161/164 (97%)

Query: 1 MAFKDNAVELEERVVAINRVTKVVGGRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIR 60
 MAFKDNAVELEERVVAINRVTKVVGGRRLRFAALVVVGD NGRVGFGTGKAQEVPEAIR
 30 Sbjct: 1 MAFKDNAVELEERVVAINRVTKVVGGRRLRFAALVVVGDNGRVGFGTGKAQEVPEAIR 60

Query: 61 KAVEAAKKNMVEVPMVGTTPHEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVA 120
 KAVEAAKKNM+EVPVPMVGTTPHEV + FGGAKVLLKPAVEG+GVAAGGAVRAVIELAGVA
 35 Sbjct: 61 KAVEAAKKNMIEVPMVGTTPHEVYTNFGGAKVLLKPAVEGSGVAAGGAVRAVIELAGVA 120

Query: 121 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA 164
 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA
 Sbjct: 121 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA 164

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2052

A DNA sequence (GBSx2163) was identified in *S.agalactiae* <SEQ ID 6339> which encodes the amino acid sequence <SEQ ID 6340>. This protein is predicted to be 50S ribosomal protein L18 (rplR). Analysis
 45 of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4488(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9465> which encodes amino acid sequence <SEQ ID 9466>
 55 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2319-

>GP:AAB06815 GB:L47971 ribosomal protein L18 [Bacillus subtilis]
Identities = 86/120 (71%), Positives = 97/120 (80%), Gaps = 2/120 (1%)

5 Query: 4 VISKPDKNKIRQKRHRVRGKLSGTADRPRNLNIFRSNTGIYAQVIDDVAGVTLASASTLD 63
+I+K KN R KRH RVR KLSGTA+RPRLN+FRSN IYAQ+IDDV GVTLASASTLD
Sbjct: 1 MITKTSKNAARLKRHARVRRAKLSGTAERPRLNVFRSNKHIYAQIIDDVNGVTLASASTLD 60

10 Query: 64 KE--VSNGTKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALADSARENLKLF 121
K+ V + T A VG+LVA+RA KGIS+VVFDRGGYLYHGRVKALAD+ARE GLKF
Sbjct: 61 KDLNVESTGDTSAATKVGEVLAKRAAEKGISDVVFDRGGYLYHGRVKALADAAREAGLKF 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6341> which encodes the amino acid sequence <SEQ ID 6342>. Analysis of this protein sequence reveals the following:

15 Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4488(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 116/121 (95%), Positives = 120/121 (98%)

25 Query: 1 MKIVISKPDKNKIRQKRHRVRGKLSGTADRPRNLNIFRSNTGIYAQVIDDVAGVTLASAS 60
+KIVISKPDKNKIRQKRHRVRGKLSGTADRPRLN+FRSNTGIYAQVIDDVAGVTLASAS
Sbjct: 1 VKIVISKPDKNKIRQKRHRVRGKLSGTADRPRNLNVFRSNTGIYAQVIDDVAGVTLASAS 60

30 Query: 61 TLDKEVSNNGTKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALADSARENLKLF 121
TLDK+VS GTKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALAD+ARENLKLF
Sbjct: 61 TLDKDVSKGKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALADAARENLKLF 121

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 2053

A DNA sequence (GBSx2164) was identified in *S.galactiae* <SEQ ID 6343> which encodes the amino acid sequence <SEQ ID 6344>. Analysis of this protein sequence reveals the following:

40 Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1530(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA22700 GB:M57622 ribosomal protein L6 [Bacillus
stearothermophilus]
Identities = 108/178 (60%), Positives = 133/178 (74%)

50 Query: 1 MSRIGNKVITLPAQVEIINKDNVTVKGPQGLTREFNKNIGITVEGTEVTVTRPDSKE 60
M R+G K I +PAGV + N VTVKGPQG+LTR F+ ++ ITVEG +TVTRP+D K
Sbjct: 1 MXRVGKKPIEIPAGVTVTVNGNTVTVKGPQGLTTRTFHPDMTITVEGNVITVTRPSDEKH 60

55 Query: 61 MKTIHGTTTRANLNNMVVGVSSEGFKKALEMRGVGYRAQLQGSKLVLVSVGKSHQDEVEAPEG 120
+ +HGTTTR+ L NMV GVS+G++KALE+ GVGYRA QG KLVLSVG SH E+E EG
Sbjct: 61 HRALHGTTTRSLLANMVEGVSKGYEKALELVGVGYRASKQGGKLVLSVGYSHPVETIEPEEG 120

Query: 121 VTFFVPTPTTINVIGINKESVGQTAAAYVRSLSRSPPEPYKKGIRYVGEFVRRKEGKTGK 178

-2320-

+ EVP+ T I V G +K+ VG+ AA +R++R PEPYKKGIRY GE VR KEGKTGK
 Sbjct: 121 LEIEVPSQTKIIVKGADKQVRGELAAANIRAVRPPEPYKKGIRYEGELVRLKEGKTGK 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6345> which encodes the amino acid
 5 sequence <SEQ ID 6346>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1704(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 153/178 (85%), Positives = 166/178 (92%)
 Query: 1 MSRIGNKVITLPAGVEIINKDNVVTVKGPKGQLTREFNKNIGITVEGTEVTVTRENDKE 60
 MSRIGNKVIT+PAGVE+ N +NV+TVKGPKG+LTREFNKN I VEGTE+TV RENDKE
 20 Sbjct: 1 MSRIGNKVITMPAGVELTNNNNVITVKGPKGELTREFNKNIEIKVEGTEITVVRNDKE 60
 Query: 61 MKTIHGTTTRANLNNMVGVSEGFKKALEMRGVGYRAQLQGSKLVLSVGKSHQDEVEAPEG 120
 MKTIHGTTTRANLNNMVGVSEGFKK LEM+GVGYRAQLQG+KLVLSVGKSHQDEVEAPEG
 Sbjct: 61 MKTIHGTTTRANLNNMVGVSEGFKKLEMKGVGYRAQLQGTKLVL SVGKSHQDEVEAPEG 120
 25 Query: 121 VTFEVPTPTTINVIGINKESVQTAAYVRSLSPEPYKKGIRYVGEFVRKEGKTGK 178
 +TF V PT+I+V GINKE VQTAAY+RSLRSPPEPYKKGIRYVGE+VR KEGKTGK
 Sbjct: 121 ITFTVANPTSISVEGINKENVQTAAYIRSLRSPPEPYKKGIRYVGEYVRLKEGKTGK 178

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 2054

A DNA sequence (GBSx2165) was identified in *S.agalactiae* <SEQ ID 6347> which encodes the amino
 acid sequence <SEQ ID 6348>. This protein is predicted to be 30S ribosomal protein S8 (rpsH). Analysis of
 this protein sequence reveals the following:

35 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.4356(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAB06813 GB:L47971 ribosomal protein S8 [Bacillus subtilis]
 Identities = 100/132 (75%), Positives = 116/132 (87%)
 Query: 1 MVMTDPIADFLTRIRNANQAKHEVLEVPASNIKKGIADILKREGFVKNVIEEDDKQGII 60
 MVMTDPIAD LTRIRNAN +HE LE+PAS +K+ IA+ILKREGF+++VE +ED KQGII
 50 Sbjct: 1 MVMTDPIADMLTRIRNANMVRHEKLEIPASKLKREIAEILKREGFIRDVEFVEDSKQGII 60
 Query: 61 RVFLKYQNGERVITNLKRISKPLRVYTKHEDMPKVLNGLGIAIVSTSEGLLTDKEARQ 120
 RVFLKYQN ERVIT LKRISKPLRVY K ++P+VLNGLGIAI+STS+G+LTDKEAR
 Sbjct: 61 RVFLKYQNNERVITGLKRISKPLRVYAKSNEVPRVLNGLGIAIISTSQGVLTDEARA 120
 55 Query: 121 KNIGGEVLAYIW 132
 K GGEVLAY+W
 Sbjct: 121 KQAGGEVLAYVW 132

-2321-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6349> which encodes the amino acid sequence <SEQ ID 6350>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4327(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 122/132 (92%), Positives = 129/132 (97%)

Query: 1 MVMTDFIADFLTRIRNANQAKHEVLEVPASNIKKGIADILKREGFVKNVIEDDKQGII 60

MVMTDFIADFLTRIRNANQ KHEVLEVPASNIKKGIA+ILKREGFVKNVIEDDKQGII

Sbjct: 1 MVMTDFIADFLTRIRNANQVKHEVLEVPASNIKKGIAEILKREGFVKNVIEDDKQGII 60

Query: 61 RVFLKYGQNGERVITNLKRISKPGRLVYTKHEDMPKVLNGLGIAIVSTSEGLLTDKEARQ 120

RVFLKYG+NGERVITNLKRISKPGRLVY K +DMPKVLNGLGIAI+STSEGLLTDKEARQ

Sbjct: 61 RVFLKYGKNGERVITNLKRISKPGRLVYAKRDDMPKVLNGLGIAIISTSEGLLTDKEARQ 120

Query: 121 KNIGGEVLAYIW 132

KN+GGEV+AY+W

Sbjct: 121 KNVGGEVIAYVW 132

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2055

A DNA sequence (GBSx2166) was identified in *S.agalactiae* <SEQ ID 6351> which encodes the amino acid sequence <SEQ ID 6352>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11905 GB:Z99104 ribosomal protein S14 [Bacillus subtilis]

Identities = 47/61 (77%), Positives = 53/61 (86%)

Query: 1 MAKKSMTAKNRPAPKSTQAYTRCEKCGRPHSVYRKQFQLCRVCFRDLAYKGQVPGVTKAS 60

MAKKSMTAK +R KF Q YTRCE+CGRPHSV RKF+LCR+CFR+LAYKGQ+PGV KAS

Sbjct: 1 MAKKSMTAKQQRTPKFKVQBYTRCERCGRPHSVIRKFKLCRICFRELAYKGQIPGVTKAS 60

Query: 61 W 61

W

Sbjct: 61 W 61

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6353> which encodes the amino acid sequence <SEQ ID 6354>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2322-

bacterial cytoplasm --- Certainty=0.4747(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 55/61 (90%), Positives = 59/61 (96%)

Query: 1 MAKKSMIANKRPKAFSTQAYTRCEKCGRPHSVYRKFKLCRVCFRDLAYKGQVPGVTKAS 60
 +AKKSMIANKRKPAK STQAYTRCEKCGRPHSVYRKFKLCRVCFR+LAYKGQ+PGV KAS
 10 Sbjct: 1 LAKKSMIANKRPKAKHSTQAYTRCEKCGRPHSVYRKFKLCRVCFRELAYKGQIPGVVKAS 60

Query: 61 W 61
 W
 Sbjct: 61 W 61

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2056

A DNA sequence (GBSx2167) was identified in *S.galactiae* <SEQ ID 6355> which encodes the amino acid sequence <SEQ ID 6356>. This protein is predicted to be 50S ribosomal protein L5 (rplE). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1845(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB03865 GB:AP001507 ribosomal protein L5 (BL6) [Bacillus halodurans]
 Identities = 143/178 (80%), Positives = 162/178 (90%)

35 Query: 3 NRLKEKYTNEVVPALTEKFNYSSVMAVPKVEKIVLNMVGVDVAVSNAKLEKAAELALIS 62
 NRLKEKY E+VP+LTEKFNYSSVMAVPK+EKIV+NMVGVDVAV NAK L+KA EL I+
 Sbjct: 2 NRLKEKYQKEIVPSLTEKFNYSSVMAVPKLEKIVNMVGVDVAVQNAKALDKAVEELTEIT 61

Query: 63 GQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVPTKS 122
 GQKP+ITKAKKSIAGF+LREG+ IGAKVTLRGERMYEFLDKL+SVSLPRVRDF G+ K+
 40 Sbjct: 62 GQKPIITKAKKSIAGFKLREGMPIGAKVTLRGERMYEFLDKLISVSLPRVRDFRGISKKA 121

Query: 123 FDGRGNYTLGVKEQLIFPEINFDDVDKVRGLDIVVTTANTDEESRELLKGLGMPFAK 180
 FDGRGNYTLGVKEQLIFPEI++D VDKVRG+D+VIVTTA+TDEE+RELL +GMPF K
 45 Sbjct: 122 FDGRGNYTLGVKEQLIFPEIDYDKVDKVRGMDVVIVTTASTDEEARELLSQMGMPFQK 179

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6357> which encodes the amino acid sequence <SEQ ID 6358>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1793(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 177/180 (98%), Positives = 180/180 (99%)

-2323-

Query: 1 MANRLKEKYTNEVPALTEKFNYSSVMAVPKVEKIVLNMGVGDAVSNAKNLEKAAAEALAL 60
 Sbjct: 1 MANRLKEKYTNEVPALTEKFNYTSVMAVPKVEKIVLNMGVGDAVSNAKNLEKAAAEALAL 60

Query: 61 ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSLSLPRVRDFHGVPT 120
 ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSLSLPRVRDFHGVPT
 Sbjct: 61 ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSLSLPRVRDFHGVPT 120

Query: 121 KSFDGRGNITLGVKEQLIFPEINFDDVDKVRGLDIVIVITTANTDEESRELLKGLGMPFAK 180
 KSFDGRGNITLGVKEQLIFPEI+FDVDKVRGLDIVIVITTANTDEESRELLKGLGMPFAK
 Sbjct: 121 KSFDGRGNITLGVKEQLIFPEISFDDVDKVRGLDIVIVITTANTDEESRELLKGLGMPFAK 180

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2057

A DNA sequence (GBSx2169) was identified in *S.agalactiae* <SEQ ID 6359> which encodes the amino acid sequence <SEQ ID 6360>. This protein is predicted to be 50S ribosomal protein L24 (rplX). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1850(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33285 GB:AF126061 Rpl24 [Streptococcus pneumoniae]
 Identities = 89/101 (88%), Positives = 94/101 (92%)

Query: 1 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNVVVEGVALIKKHQKPNENPQGAIVEKE 60
 MFVKKGDKVRVIAGKDKGTEAVVL ALPKVNVV+VEGV ++KKHQ+P NE PQG I+EKE
 Sbjct: 1 MFVKKGDKVRVIAGKDKGTEAVVLTALPKVNVVVEGVNIVKKHQPTNELPQGGIIEKE 60

Query: 61 APIHVSINVQVLDKNGVAGRVGYKVVDGKKVRYNKKSGEVL 101
 A IHVSINVQVLDKNGVAGRVGYK VDGKKVRYNKKSGEVL
 Sbjct: 61 AAIHVSINVQVLDKNGVAGRVGYKFVDGKKVRYNKKSGEVL 101

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6361> which encodes the amino acid sequence <SEQ ID 6362>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1850(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 95/101 (94%), Positives = 99/101 (97%)

Query: 1 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNVVVEGVALIKKHQKPNENPQGAIVEKE 60
 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNVV+VEGV +IKKHQKPN ENPQGAIVEKE
 Sbjct: 1 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNVVVEGVGMIIKKHQPTNELPQGAIVEKE 60

Query: 61 APIHVSINVQVLDKNGVAGRVGYKVVDGKKVRYNKKSGEVL 101
 APIHVSINVQVLDKNGVAGR+GYKVVDGKKVRY+KKSGEVL

-2324-

Sbjct: 61 APIHVSINVQVLDKNGVAGRIGYKVVVDGKKVRYSKKSGEVLD 101

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2058

A DNA sequence (GBSx2170) was identified in *S.agalactiae* <SEQ ID 6363> which encodes the amino acid sequence <SEQ ID 6364>. This protein is predicted to be 50S ribosomal protein L14 (rplN). Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1004(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33284 GB:AF126061 RplL14 [Streptococcus pneumoniae]
 Identities = 116/122 (95%), Positives = 120/122 (98%)

Query: 1 MIQQETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60
 MIQ ETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA
 Sbjct: 1 MIQTETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60

Query: 61 VIVRTKTGARRPDGSYIKFDDNAAVIIRD DKT PRGTRIFGPVARELREGGYMKIVSLAPE 120
 VIVRTK+GARR DGSYIKFD+NAAVLIIR+DKTPRGTRIFGPVARELREGG+MKIVSLAPE
 Sbjct: 61 VIVRTKSGARRADGSYIKFDENAAVIIRD DKT PRGTRIFGPVARELREGGFMKIVSLAPE 120

Query: 121 VL 122
 VL
 Sbjct: 121 VL 122

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6365> which encodes the amino acid sequence <SEQ ID 6366>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1004(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 122/122 (100%), Positives = 122/122 (100%)

Query: 1 MIQQETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60
 MIQQETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA
 Sbjct: 1 MIQTETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60

Query: 61 VIVRTKTGARRPDGSYIKFDDNAAVIIRD DKT PRGTRIFGPVARELREGGYMKIVSLAPE 120
 VIVRTKTGARRPDGSYIKFDDNAAVIIRD DKT PRGTRIFGPVARELREGGYMKIVSLAPE
 Sbjct: 61 VIVRTKTGARRPDGSYIKFDDNAAVIIRD DKT PRGTRIFGPVARELREGGYMKIVSLAPE 120

Query: 121 VL 122
 VL
 Sbjct: 121 VL 122

-2325-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2059

5 A DNA sequence (GBSx2171) was identified in *S.agalactiae* <SEQ ID 6367> which encodes the amino acid sequence <SEQ ID 6368>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3415(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAD33283 GB:AF126061 RpS17 [Streptococcus pneumoniae]
Identities = 82/86 (95%), Positives = 83/86 (96%)

Query: 1 MERNQRTLYGRVSDKMDKTIITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60
MERN RK L GRVSDKMDKTIITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV

20 Sbjct: 1 MERNNRKVLVGRVSDKMDKTIITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRPLSATKRFRLEVEVEKAVII 86
RIMETRPLSATKRFRLEVEVE+AVII

25 Sbjct: 61 RIMETRPLSATKRFRLEVEVEEAVII 86

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6369> which encodes the amino acid sequence <SEQ ID 6370>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3415(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/86 (100%), Positives = 86/86 (100%)

40 Query: 1 MERNQRTLYGRVSDKMDKTIITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60
MERNQRTLYGRVSDKMDKTIITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV
Sbjct: 1 MERNQRTLYGRVSDKMDKTIITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRPLSATKRFRLEVEVEKAVII 86
RIMETRPLSATKRFRLEVEVEKAVII

45 Sbjct: 61 RIMETRPLSATKRFRLEVEVEKAVII 86

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2060

50 A DNA sequence (GBSx2172) was identified in *S.agalactiae* <SEQ ID 6371> which encodes the amino acid sequence <SEQ ID 6372>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

-2326-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4329(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33282 GB:AF126061 Rpl29 [Streptococcus pneumoniae]
 Identities = 58/68 (85%), Positives = 64/68 (93%)

10 Query: 1 MKLQEIKDFVKELRGLSQEELAKKENELKKELFDLRFQAAAGQLEKTARLDEVKKQIARV 60
 MKL E+K+FKELRGLSQEELAK+ENELKKELF+LRFQAA GQLE+TARL EVKKQIAR+
 Sbjct: 1 MKLNEVKEFVKELRGLSQEELAKRENELKKELFELRFQAAATGQLEQTARLKEVKKQIARI 60

15 Query: 61 KTVQSEMK 68
 KTVQSE K
 Sbjct: 61 KTVQSEAK 68

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 **Example 2061**

A DNA sequence (GBSx2174) was identified in *S.agalactiae* <SEQ ID 6373> which encodes the amino acid sequence <SEQ ID 6374>. This protein is predicted to be Rpl16 (rplP). Analysis of this protein sequence reveals the following:

25 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.4574(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33263 GB:AF126059 Rpl16 [Streptococcus pneumoniae]
 Identities = 135/137 (98%), Positives = 137/137 (99%)

35 Query: 1 MLVPKRVKVRREFRFGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60
 MLVPKRVKVRREFRFGKMRGEAKGGKEV+FGYGLQATTSHWITNRQIEAARIAMTRYMKR
 Sbjct: 1 MLVPKRVKVRREFRFGKMRGEAKGGKEVAFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60

40 Query: 61 GGVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEEVAREALRL 120
 GGVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEE+AREALRL
 Sbjct: 61 GGVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEEIAREALRL 120

45 Query: 121 ASHKLPVKCKFVKREAE 137
 ASHKLPVKCKFVKREAE
 Sbjct: 121 ASHKLPVKCKFVKREAE 137

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6375> which encodes the amino acid sequence <SEQ ID 6376>. Analysis of this protein sequence reveals the following:

50 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.4574(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2327-

Identities = 136/137 (99%), Positives = 137/137 (99%)

Query: 1 MLVPKRVKVRHREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60
 5 Sbjct: 1 MLVPKRVKVRHREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60

Query: 61 GGVVWIKIFPHKSYTAKAIGVRMGSGKGAPGEGWVAPVIRGKVMFEIAGVSEEVAREALRL 120
 10 Sbjct: 61 GGVVWIKIFPHKSYTAKAIGVRMGSGKGAPGEGWVAPVIRGKVMFEIAGVSEE+AREALRL 120

Query: 121 ASHKLFVKCKFVKREAE 137
 ASHKLFVKCKFVKREAE
 Sbjct: 121 ASHKLFVKCKFVKREAE 137

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2062

A DNA sequence (GBSx2175) was identified in *S.galactiae* <SEQ ID 6377> which encodes the amino acid sequence <SEQ ID 6378>. Analysis of this protein sequence reveals the following:

20 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.3758(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAD33280 GB:AF126061 RpS3 [Streptococcus pneumoniae]
 Identities = 200/208 (96%), Positives = 203/208 (97%)

Query: 10 MRVGIIRDWDKWAKEYADYLHEDLAIRKPFINKELADASVTIEIERAVNKVIVSLHT 69
 MRVGIIRDWDKWAKEYADYLHEDLAIRKF+ KELADA+VSTIEIERAVNKV VSLHT
 35 Sbjct: 1 MRVGIIRDWDKWAKEYADYLHEDLAIRKFVQKELADAAVSTIEIERAVNKVIVSLHT 60

Query: 70 AKPGMVIGKGGANVDALRGQLNKLTKGQVHINIIEIKQPDLDLHVLGENIARQLEQRVAF 129
 AKPGMVIGKGGANVDALR +LNKLTKGQVHINIIEIKQPDLDLHVLGE IARQLEQRVAF
 40 Sbjct: 61 AKPGMVIGKGGANVDALRAKLNKLTKGQVHINIIEIKQPDLDLHVLGEGIARQLEQRVAF 120

Query: 130 RRAQKQAIQRTMRAGAKGIKTQVSGRLNGADIARAEGYSEGTVPPLHTLRADIDYAWEEAD 189
 RRAQKQAIQR MRAGAKGIKTQVSGRLNGADIARAEGYSEGTVPPLHTLRADIDYAWEEAD
 45 Sbjct: 121 RRAQKQAIQRAMRAGAKGIKTQVSGRLNGADIARAEGYSEGTVPPLHTLRADIDYAWEEAD 180

Query: 190 TTYGKLGVKVWIYRGEVLPARKNTKGGK 217
 TTYGKLGVKVWIYRGEVLPARKNTKGGK
 50 Sbjct: 181 TTYGKLGVKVWIYRGEVLPARKNTKGGK 208

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6379> which encodes the amino acid sequence <SEQ ID 6380>. Analysis of this protein sequence reveals the following:

50 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.3758(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2328-

Example 2063

A DNA sequence (GBSx2176) was identified in *S.agalactiae* <SEQ ID 6381> which encodes the amino acid sequence <SEQ ID 6382>. This protein is predicted to be 50S ribosomal protein L22 (rplV). Analysis of this protein sequence reveals the following:

```

5   Possible site: 36
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2704(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAD33279 GB:AF126061 Rpl22 [Streptococcus pneumoniae]
      Identities = 99/114 (86%), Positives = 106/114 (92%)

      Query: 1   MAEITSAKAMARTVRVSPRKTRLVLDLIRGKNVADAIAILKFTPNKAARVIEKTLNSAIA 60
                MAEITSAKAMARTVRVSPRK+RLVLD IRGK+VADAIAIL FTPNKAA +I K LNSA+A
20   Sbjct: 1   MAEITSAKAMARTVRVSPRKSRLVLDNIRGKSVADAIAILTFTPNKAAEIIKVLNSAVA 60

      Query: 61  NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
                NAENNFGGL+KANLVVSE FANEGPTMKRFRPRAKGSASPINKRT H+TV V+EK
      Sbjct: 61  NAENNFGLDKANLVVSEAFANEGPTMKRFRPRAKGSASPINKRTAHITVAVAEK 114

```

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6383> which encodes the amino acid sequence <SEQ ID 6384>. Analysis of this protein sequence reveals the following:

```

      Possible site: 36
      >>> Seems to have no N-terminal signal sequence

30   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2794(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 113/114 (99%), Positives = 113/114 (99%)

      Query: 1   MAEITSAKAMARTVRVSPRKTRLVLDLIRGKNVADAIAILKFTPNKAARVIEKTLNSAIA 60
                MAEITSAKAMARTVRVSPRKTRLVLDLIRGK VADAIAILKFTPNKAARVIEKTLNSAIA
40   Sbjct: 1   MAEITSAKAMARTVRVSPRKTRLVLDLIRGKKVADAIAILKFTPNKAARVIEKTLNSAIA 60

      Query: 61  NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
                NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK
      Sbjct: 61  NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
45

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2064

50 A DNA sequence (GBSx2177) was identified in *S.agalactiae* <SEQ ID 6385> which encodes the amino acid sequence <SEQ ID 6386>. This protein is predicted to be 30S ribosomal protein S19 (rpsS). Analysis of this protein sequence reveals the following:

```

      Possible site: 23
      >>> Seems to have no N-terminal signal sequence

55   ----- Final Results -----

```

-2329-

bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein is similar to ribosomal protein S19 from *S.pneumoniae*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6387> which encodes the amino acid sequence <SEQ ID 6388>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3319(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 92/92 (100%), Positives = 92/92 (100%)

Query: 1 MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAYVDGRKHVPV 60
 MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAYVDGRKHVPV
 Sbjct: 19 MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAYVDGRKHVPV 78

Query: 61 YIQEDMVGHKLGEFAPTRTRYKGHAADKKTRR 92
 YIQEDMVGHKLGEFAPTRTRYKGHAADKKTRR

Sbjct: 79 YIQEDMVGHKLGEFAPTRTRYKGHAADKKTRR 110

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2065

30 A DNA sequence (GBSx2178) was identified in *S.agalactiae* <SEQ ID 6389> which encodes the amino acid sequence <SEQ ID 6390>. This protein is predicted to be L2 (rplB). Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3182(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45959 GB:U43929 L2 [Bacillus subtilis]
 Identities = 208/277 (75%), Positives = 239/277 (86%)

45 Query: 1 MGIVKYKPTTNGRRNMTSLDFAEITTTNTPKESLLVSLKNKAGRNNNGRITVRHQGGGHR 60
 M IK YKP++NGRR MT+ DFAEITT+ PEKSL L K GRNN G++TVRHQGGGHR
 Sbjct: 1 MAIKKYKPSSNGRRGMTTSDFAEITTDKPEKSL LAPLHKKGRRNQGLTVRHQGGGHR 60

50 Query: 61 HYRLIDFKRNDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYILAPKGLEVGQRIISGP 120
 YR+IDFKR+KDG+ V T+EYDPNR+ANIAL++Y DG K YILAPKG++VG ++SGP
 Sbjct: 61 QYRVIDFKRDKGIPGRVATVEYDPNRSANIALINADGEKRYILAPKGIQVGTVEVMSP 120

55 Query: 121 EADIKVGNALPLANIPVGTVIHNIELQPGKGAELIRAAGASAQVLGQEGKYVLVRLQSGE 180
 EADIKVGNALPL NIPVGTV+HNIEL+PGKG +L+R+AG SAQVLG+EGKYVLVRL SGE
 Sbjct: 121 EADIKVGNALPLINIPVGTVVHNIELKPGKGQLVRSAGTSAQVLGKEGKYVLVRLNSGE 180

Query: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRNRWKGVRPTVRGSMNPNDHPHGGGEGKAP 240

-2330-

VRMIL CRA+IG VGNEQ L+NIGKAGR+RWKG+RPTVRGSMNPNDHPHGGGEG+AP
 Sbjct: 181 VRMILSACRASIGQVGNEQHELINIGKAGRSRWKGIRPTVRGSMNPNDHPHGGGEGRAP 240

Query: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIIVRRRNQK 277
 +GRK+P +PWGKP LG KTR KK KSDK IIVRRR K
 Sbjct: 241 IGRKSPMSPWGKPTLGFKTRKKKNKSDKFIVRRRNKNK 277

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6391> which encodes the amino acid sequence <SEQ ID 6392>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2560(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 264/277 (95%), Positives = 276/277 (99%)

Query: 1 MGIKVYKPTTNGRRNMTSLDFAEITTTTPEKSLVSLKKNAGRNNNGRITVRHQGGGHR 60
 +GIKVKPTTNGRRNMTSLDFAEITT+TPEKSLVSLK+KAGRNNNGRITVRHQGGGHR
 Sbjct: 1 VGIKVYKPTTNGRRNMTSLDFAEITTTTPEKSLVSLKSKAGRNNNGRITVRHQGGGHR 60

Query: 61 HYRLIDFKRNKDGVEAVVKTIEYDENRTANIALVHYTDGVKAYILAPKGLEVGQRIISGP 120
 HYRLIDFKRNKDGVEAVVKTIEYDENRTANIALVHYTDGVKAYI+APKGLEVGQRI+SGP
 Sbjct: 61 HYRLIDFKRNKDGVEAVVKTIEYDENRTANIALVHYTDGVKAYIIAPKGLEVGQRIVSGP 120

Query: 121 EADIKVGNALPLANIPVGTVIHNIELQPGKGAEILRAAGASAQVLGQEGKYVLVRLQSGE 180
 +ADIKVGNALPLANIPVGTV+HNIEL+PGKG EL+RAAGASAQVLGQEGKYVLVRLQSGE
 Sbjct: 121 DADIKVGNALPLANIPVGTVVHNIELKPGKGELVRAAGASAQVLGQEGKYVLVRLQSGE 180

Query: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRNKRWGVRPTVRGSMNPNDHPHGGGEGKAP 240
 VRMILGTCRATIGTVGNEQQSLVNIGKAGR+RWKG+RPTVRGSMNPNDHPHGGGEGKAP
 Sbjct: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRSRWKGIRPTVRGSMNPNDHPHGGGEGKAP 240

Query: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIIVRRRNQK 277
 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIIVRRRN+K
 Sbjct: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIIVRRRNEK 277

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2066

A DNA sequence (GBSx2180) was identified in *S.agalactiae* <SEQ ID 6393> which encodes the amino acid sequence <SEQ ID 6394>. This protein is predicted to be 50S ribosomal protein L23 (rplW). Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1669(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB03855 GB:AP001507 ribosomal protein L23 [Bacillus halodurans]
 Identities = 56/92 (60%), Positives = 67/92 (71%), Gaps = 1/92 (1%)

-2331-

Query: 2 NLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVTVKP 61
 N DVIK+PVITE+S + KYTFEVD RA+K IK A+E FD VKVA VNT+ K
 Sbjct: 3 NARDVIKRPVITERSTEVMDKKYTFEVDVRANKTQIKDAIEEIFD-VKVAKVNTMNYKG 61

5 Query: 62 KAKRVGRYTGFTSKTKKAIITLTADSKAIELF 93
 K KR GRYTGFT++ KKAI+TLT DSK ++ F
 Sbjct: 62 KPKRFGRYTGFTARRKKAIIVTLTPDSKELDF 93

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6395> which encodes the amino acid
 10 sequence <SEQ ID 6396>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1617(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 96/98 (97%), Positives = 97/98 (98%)
 Query: 1 MNLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVTVK 60
 MNLYDVIKKPVITEKSM+ALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTV VK
 25 Sbjct: 1 MNLYDVIKKPVITEKSMIALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVNVK 60
 Query: 61 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98
 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE
 Sbjct: 61 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 2067

A DNA sequence (GBSx2181) was identified in *S.agalactiae* <SEQ ID 6397> which encodes the amino
 acid sequence <SEQ ID 6398>. This protein is predicted to be 50S ribosomal protein L4 (rpL4). Analysis of
 35 this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.54 Transmembrane 140 - 156 (139 - 156)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.1617(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45957 GB:U43929 L4 [Bacillus subtilis]
 Identities = 130/207 (62%), Positives = 160/207 (76%)

50 Query: 1 MANVKLFQDTGKEVSSVELNEAIFGIEPNESVVFDDVISQRASLRQGTTHAVKNRSVAVSGG 60
 M V L++Q G +ELN ++FGIEPNESVVFDD ++ QRASLRQGTTH VKNRS V GG
 Sbjct: 1 MPKVALYNQNGSTAGDIELNASVFGIEPNESVVFDAIIMQRASLRQGTTHKVNRSVAVSGG 60

Query: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSAKVAE 120
 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSY YKLP+KVRRLA+KSV S+KV +
 55 Sbjct: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPKVRRLA+KSVLSSKVID 120

Query: 121 DKFVAVENLSFAAPKTAEFASVLSALSIDSKVLVILEEGNEFAALSARNLPNVTATATT 180
 + + +E+L+ KT E A++L Ls++ K L++ + NE ALSARN+P VTV A

-2332-

Sbjct: 121 NNIIVLEDLTLDTAKTEMAAILKGLSVEKKALIVTADANEAVALSARNIPGVTVEANG 180

Query: 181 ASVLDIVNADKLLVTKEAISTIEGVLA 207
+VLD+VN +KLL+TK A+ +E VLA

5 Sbjct: 181 INVLDVVNHEKLLITKAAVEKVEEVLA 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6399> which encodes the amino acid sequence <SEQ ID 6400>. Analysis of this protein sequence reveals the following:

Possible site: 60

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2544(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 199/207 (96%), Positives = 203/207 (97%)

20 Query: 1 MANVKLFDQTGKEVSSVELNEAIFGIEPNESVVFDDVISQRASLRQGTHAVKNRSVAVSGG 60
MANVKLFDQTGKEVSSVELN+AIFGIEPNESVVFDDVISQRASLRQGTHAVKNRSVAVSGG
Sbjct: 1 MANVKLFDQTGKEVSSVELNDAIFGIEPNESVVFDDVISQRASLRQGTHAVKNRSVAVSGG 60

25 Query: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLQKVRRLALKSVYSAKVAE 120
GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLQKVRRLALKSVYSAKVAE
Sbjct: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLQKVRRLALKSVYSAKVAE 120

30 Query: 121 DKFVAVENLSFAAPKTAEFASVLSALSIDSKVLVILEEGNEFAALSARNLPNVTVATATT 180
DKFVAVE LSFAAPKTAEFA VLSALSID+KVLV++EEGNEFAALSARNLPNVTVATA T
Sbjct: 121 DKFVAVEGLSFAAPKTAEFKVLVLSALSIDTKVLVLEEGNEFAALSARNLPNVTVATAAT 180

Query: 181 ASVLDIVNADKLLVTKEAISTIEGVLA 207
ASVLDIVNADKLLVTKEAISTIE VLA

35 Sbjct: 181 ASVLDIVNADKLLVTKEAISTIEEVLA 207

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2068

40 A DNA sequence (GBSx2183) was identified in *S.agalactiae* <SEQ ID 6401> which encodes the amino acid sequence <SEQ ID 6402>. This protein is predicted to be 50S ribosomal protein L3 (rplC). Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2090(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45956 GB:U43929 L3 [Bacillus subtilis]

Identities = 157/208 (75%), Positives = 180/208 (86%), Gaps = 2/208 (0%)

55 Query: 1 MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKTVEIDGYEAVQVGFDDKREVL 60
MTKGILG+K+GMTQ+F E+G+ IPVTVIEA PNVVLQ KT E DGYEA+Q+GFDDKRE L
Sbjct: 1 MTKGILGRKIGMTQVFAENGDLIPVTVIEAAPNVVLQKKTAEENDGYEAIQLGFDDKREKL 60

Query: 61 SNKPAKGHVAKANTAPKRFIREFKNIE--GLEVGAELSVEQFEAGDVVDVTGTSKGKGFGQ 118

-2333-

SNKP KGHVAKA TAPKRF++E + +E EVG E+ VE F AG++VDVTG SKGKGFGQ
 Sbjct: 61 SNKPEKGHVAKAETAPKRFVKELRGVEMDAYEVGQEVKVEIFSAGEIVDVTGVSCKGKGFGQ 120

Query: 119 GVIKRHGQSRGPMAGHSRYHRRPGSMGPVAPNRFVFNKRLAGRMGGNRVTQNLIVQVI 178
 5 G IKRHGQSRGPM+HGSRYHRRPGSMGPV PNRVFK K L GRMGG ++TVQNLEIV+V
 Sbjct: 121 GAIKRHGQSRGPMHGSRYHRRPGSMGPVDPNRVFKGLLPGRMGGEQITVQNLEIVKVD 180

Query: 179 PEKNVVLKGNVPGAKKSLITIKSAVKA 206
 E+N++LIKGNVPGAKKSLIT+KSAVK+
 10 Sbjct: 181 AERNLLLIKGNVPGAKKSLITVKS AVKS 208

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6403> which encodes the amino acid sequence <SEQ ID 6404>. Analysis of this protein sequence reveals the following:

Possible site: 40
 15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2090(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 205/208 (98%), Positives = 207/208 (98%)

Query: 1 MTKGILGKKVGMTQIFTESGEFIPVTVEATPNVVLQVKTVDGYEAVQVGFDDKREVL 60
 MTKGILGKKVGMTQIFTESGEFIPVTVEATPNVVLQVKTVDGYEAVQVGFDDKREVL
 Sbjct: 1 MTKGILGKKVGMTQIFTESGEFIPVTVEATPNVVLQVKTVDGYEAVQVGFDDKREVL 60

Query: 61 SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAELSVEQFEAGDVVDVTGTSKKGFGQGV 120
 SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAELSVEQFEAGDVVDVTG SKGKGFGQGV
 30 Sbjct: 61 SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAELSVEQFEAGDVVDVTGISKKGFGQGV 120

Query: 121 IKRHGQSRGPMAGHSRYHRRPGSMGPVAPNRFVFNKRLAGRMGGNRVTQNLIVQVIPE 180
 IKRHGQSRGPMAGHSRYHRRPGSMGPVAPNRFVFNKRLAGRMGGNRVTQNLIVQVIPE
 35 Sbjct: 121 IKRHGQSRGPMAGHSRYHRRPGSMGPVAPNRFVFNKRLAGRMGGNRVTQNLIVQVIPE 180

Query: 181 KNVVLKGNVPGAKKSLITIKSAVKA 208
 KNV+L+KGNVPGAKKSLITIKSAVKA
 40 Sbjct: 181 KNVILVKGNVPGAKKSLITIKSAVKA 208

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2069

A DNA sequence (GBSx2184) was identified in *S.agalactiae* <SEQ ID 6405> which encodes the amino acid sequence <SEQ ID 6406>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.43 Transmembrane 5 - 21 (5 - 21)

----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2334-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2070

A DNA sequence (GBSx2185) was identified in *S.agalactiae* <SEQ ID 6407> which encodes the amino acid sequence <SEQ ID 6408>. This protein is predicted to be 30S ribosomal protein S10 (rpsJ). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3160(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB46363 GB:L29637 S10 ribosomal protein [Streptococcus mutans]
Identities = 98/102 (96%), Positives = 102/102 (99%)

Query: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD 60
MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGA+VAGPVPLPTERSLYT+IRATHKYKD
Sbjct: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGASVAGPVPLPTERSLYTVIRATHKYKD 60

Query: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102
SREQFEMRTHKRL+DI+NPTQKTVDALMKLDLPSGVNVEIKL
Sbjct: 61 SREQFEMRTHKRLDIVNPTQKTVDALMKLDLPSGVNVEIKL 102

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6409> which encodes the amino acid sequence <SEQ ID 6410>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.3160(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 102/102 (100%), Positives = 102/102 (100%)
Query: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD 60
MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD
Sbjct: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD 60
Query: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102
SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL
Sbjct: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2071

A DNA sequence (GBSx2186) was identified in *S.agalactiae* <SEQ ID 6411> which encodes the amino acid sequence <SEQ ID 6412>. Analysis of this protein sequence reveals the following:

Possible site: 34

-2335-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2538(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2072

A DNA sequence (GBSx2187) was identified in *S.agalactiae* <SEQ ID 6413> which encodes the amino acid sequence <SEQ ID 6414>. Analysis of this protein sequence reveals the following:

15 Possible site: 46

>>> Seems to have no N-terminal signal sequence

20 INTEGRAL Likelihood = -11.41 Transmembrane 88 - 104 (79 - 110)
 INTEGRAL Likelihood = -8.39 Transmembrane 304 - 320 (300 - 324)
 INTEGRAL Likelihood = -6.58 Transmembrane 185 - 201 (180 - 206)
 INTEGRAL Likelihood = -5.63 Transmembrane 338 - 354 (331 - 357)
 INTEGRAL Likelihood = -5.52 Transmembrane 240 - 256 (237 - 259)
 INTEGRAL Likelihood = -4.99 Transmembrane 383 - 399 (375 - 407)
 INTEGRAL Likelihood = -3.82 Transmembrane 49 - 65 (48 - 73)
 INTEGRAL Likelihood = -2.87 Transmembrane 127 - 143 (121 - 144)
 INTEGRAL Likelihood = -2.81 Transmembrane 159 - 175 (159 - 177)
 INTEGRAL Likelihood = -2.18 Transmembrane 30 - 46 (30 - 47)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:BAB06655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 132/423 (31%), Positives = 210/423 (49%), Gaps = 16/423 (3%)

40 Query: 7 IIQLAIPAMIEINILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITIYQAIF--IALGASI 64
 + L P IE +L MLMG D +++Q AV+ V V+N I+ + +F +A G SI
 Sbjct: 11 LFALTWPFIIEILLHMLMGNADTILMSQYSDDAVAAGVSNQILAVIIVMFGFVATGTSTI 70

45 Query: 65 ASLLAKSLAGSKDDAISVCSQAIFLTLLIGAVLGIISIVFGQTFFKLLGTTKSVAVQVGG 124
 L+A+ L ++++A V +I L+ G VLG++ I FG K + S+ Q
 Sbjct: 71 --LVAQHLGAKERENAGKVAVVSIGANLIFGIVLGLLLIAFGPPILKAMQLDDSLLEAT 128

50 Query: 125 LYLAIVGGGVVTLGMLTTLGSFLRVQGPRLPMYVSIFVNFLNAVLSGFATFEWR----Y 180
 LYL IVGG V ++ T G+ LR + MYV+I +N LN + + IF
 Sbjct: 129 LYLQIVGGFSVVQSLIMTAGAILRSHSFTKDVMYVTIGMNLNVIGNYLFIFGPFIPVL 188

55 Query: 181 GLVGVAVSTLIARLIGICILAKYL-----PIKKIIKRMTWKISAQIWNLALPSAGER 232
 G+ GVA+ST+++R IG+ ++A L P ++KR + + +PSAGE+
 Sbjct: 189 GVTGVALSTVVSRTIGLEFVIAILLYKRIGELPFAYLLKRFPRVELRNLLKIGIPSAGEQ 248

60 Query: 233 LMMRAGDVVIVAIVVQLGTNNVAGNAIGETLTQFNYPGLGIATATIILTAKYVGQKNRE 292
 L A +VI + +GT + + L F ++ + I T IL VG K +
 Sbjct: 249 LSYNASQLVITYFIAMMGTEALTTKVYTQNLMMFVFLFAVAIGQGTQILIGHQVGAKIQ 308

65 Query: 293 SIEETIQSSYYIGLVLMILISSFMLLAGKPLTQLFTNNPSAIGSLIVILLSFVGVPATI 352
 + S +I + + + ++ PL +FT+NP + ++LL+ + P
 Sbjct: 309 AAYVRCFRSLWIAMTVSVSMVAVFFAFSTPLLGIPTDNPDIILSLGTTLLLLTIILEPGR 368

-2336-

Query: 353 GTLVYTAAWQGLGNAKLPFYTTTIGMWLIRVVLGYLLGIVFELGLLGVMATADNIFRW 412
 LV ++ + G+ K P Y + MW I V + YLLG+ LGL+GVW+A IAD FR
 Sbjct: 369 CNLVVSSSLRAAGDVKFPVYLAIVSMWGIAPVPIAYLLGLPLGLGLIGVWIAFIADWFRG 428

Query: 413 LFL 415
 L +
 Sbjct: 429 LLM 431

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6415> which encodes the amino acid sequence <SEQ ID 6416>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

15	INTEGRAL	Likelihood = -5.26	Transmembrane	89 - 105 (85 - 108)
	INTEGRAL	Likelihood = -4.35	Transmembrane	305 - 321 (302 - 322)
	INTEGRAL	Likelihood = -3.82	Transmembrane	161 - 177 (161 - 180)
	INTEGRAL	Likelihood = -3.82	Transmembrane	192 - 208 (189 - 208)
	INTEGRAL	Likelihood = -3.77	Transmembrane	129 - 145 (128 - 151)
	INTEGRAL	Likelihood = -3.24	Transmembrane	242 - 258 (240 - 258)
20	INTEGRAL	Likelihood = -2.81	Transmembrane	378 - 394 (377 - 394)
	INTEGRAL	Likelihood = -2.66	Transmembrane	339 - 355 (338 - 358)
	INTEGRAL	Likelihood = -2.60	Transmembrane	58 - 74 (58 - 75)
	INTEGRAL	Likelihood = -2.50	Transmembrane	32 - 48 (32 - 49)

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.3102(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 30 The protein has homology with the following sequences in the databases:

>GP:BAB06655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 119/435 (27%), Positives = 214/435 (48%), Gaps = 14/435 (3%)

Query: 9 IFSLALPSMIENILQMLMGMVDNYLVAQIGLVAVSGVSIANNIISIYQSLFIALGAAVSS 68
 +F+L P IE +L MLMG D +++Q AV+ V ++N I+++ +F + S
 Sbjct: 11 LFALTWPIFIEILLHMLMGNADTLMLSQYSDDAVAAGVSNQILAVIIVMFGFVATGTSI 70

Query: 69 LIARSIGENNQNKQLNYMAGVLQVTLILLSVGLGLLSVAGHHQVLEWLGAESVTLVGGQY 128
 L+A+ +G + + L+ + LGLL +A +L+ + + S+ Y
 Sbjct: 71 LVAQHLGAKERENAGKVAVVSIGANLIFGIVLGLLLIAFGPPILKAMQLDSSLQEQATLY 130

Query: 129 LSIIVGGMIVSLGLLTSLGAIVRAQGYPKIPMQVSLINVLNNAIFSALSIY----VWGFGL 184
 L IVGG V L+ + GAI+R+ + K M V++ +N+LN I + L I+ + G+
 Sbjct: 131 LQIVGGSFVVQSLIMTAGAILRSHSFTKDVMYVTIGMNLNVIGNYLFIFGPFGIPVLGV 190

Query: 185 LGVANATVLSRLVGVFLLCQF-----IPIKQVAKRLMRPLDKIIFDLSLPAAGERLM 236
 GVA +TV+SR +G+F++ +P + KR R + + + +P+AGE+L
 Sbjct: 191 TGVALSTVVSRTIGLFLVIAILLYKRIRGELPFAYLLKRFPRVELRNLLKIGIPSAGEQLS 250

Query: 237 MRAGDVLIIIGIVVRFGTALAGNAIGETLTQFNYPGLAMATATIILVARQLGGGKVTEI 296
 A ++I + GT AL + L F ++ +A+ T IL+ Q+G ++
 Sbjct: 251 YNASQLVITYFIAMMGTEALTTKVYVTQNLMMFVFLFAVAIGQGTQILIGHQVGAKQIQAA 310

Query: 297 RYIIREAFILSTLMMLVMGALTLYLLGPSLLPLFTQNTDAQRSAMIVLLFSLLGAPATAGT 356
 + ++ + + M + + LL +FT N D +LL +++ P A
 Sbjct: 311 YVRCFRSLWIAMTVSVSMVAVFFAFSTPLLGIFTDNPDLISLGTTLTLLLTIIIEPGRACN 370

Query: 357 LVYTAVWQGLGKAKLPFYATTIGMWVIRIGLVYIGVWVQYGLIGVWMAITVLDNTRWFI 416
 LV + + G K P Y + MW I + + Y++G+ GLIGVW+A + D R +
 Sbjct: 371 LVVSSSLRAAGDVKFPVYLAIVSMWGIAPVPIAYLLGLPLGLGLIGVWIAFIADWFRGLL 430

Query: 417 LSKHFK--KYQEITF 429
 + ++ K+QE++F
 Sbjct: 431 MIWRWRKKGWQEMSF 445

65

-2337-

An alignment of the GAS and GBS proteins is shown below.

Identities = 219/418 (52%), Positives = 316/418 (75%)

```

5  Query: 5  KEIIQLAIPAMIENTILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITTYQAIFIALGASI 64
    ++I LA+P+MIENILQMLMG+VDNYLVAQ+G+VAVSGVS+ANNII+IYQ++FIALGA++
    Sbjct: 7  RKIFSLALPSMIENILQMLMGVVDNYLVAQIGLVAVSGVSIANNIISIYQSLFIALGAAV 66

    Query: 65  ASLLAKSLAGSKKDDAISVCSQAIFLTLLIGAVLGIISIVFGQTFKLLGTTKSVAQVGG 124
    +SL+A+S+ + ++ ++ + + +TLL+ LG++S+ + LG SV VGG
10  Sbjct: 67  SSLIARSIGENNQNKQLNYMAGVLQVTLTLLSVGLGLLSVAGHHQVLEWLGAESVTLVGG 126

    Query: 125  LYLAIVGGGVVTLGMLTTLGSPLRVQCQPRLPYVSIFVFNFLNAVLSCFAIFEWRYGLVG 184
    YL+IVGG +V+LG+LT+LG+ +R QG P++PM VS+ +N LNA+ S +I+ W +GL+G
15  Sbjct: 127  QYLSIVGGMIVSLGLLTSLGAIVRAQGYPKIPMQVSLINVLNNAIFSALSIYVWGFGLLG 186

    Query: 185  VAVSTLIARLIGICILAKYLPKIKIIRMTWKISAQIWNALPSAGERLMMRAGDVVIVA 244
    VA +T+++RL+G+ +L +++PIK++ KR+ + I++L+LP+AGERLMMRAGDV+I+
20  Sbjct: 187  VAWATVLSRLVGVFLLCQFIPIKQVAKRLMRPLDKIIFDLSLPAAGERLMMRAGDVLIIG 246

    Query: 245  IVVQLGTNVVAGNAIGETLTQFNYPGLGIATATIILTAKYVGQKNRESIEETIQSSYYI 304
    IVV+ GT +AGNAIGETLTQFNYPGL +ATATTIL A+ +G I I+ ++ +
25  Sbjct: 247  IVVRFGTTALAGNAIGETLTQFNYPGLAMATATIILVARQLGGGKVTEIRYIIREAFIL 306

    Query: 305  GLVLMILISSFMLLAGKPLTQLFTNNPSAIKGLSLIVILLSFVGVDPATIGTLVYTAAWQGL 364
    ++M+++ + L G L LFT N A + ++IV+L S +G PAT GTLVYTA WQGL
30  Sbjct: 307  STLMLVMGALTLYLLGPSLLPLFTQNTDAQRSAMIVLLFSLLGAPATAGTLVYTAVWQGL 366

    Query: 365  GNAKLPHYTTTIGMMLIRVVLGYLLGIVFELGGLGVWMATIADNIIFRWLFLKVHYHRY 422
    G AKLPHY TTIGMW+IR+ LGY++G+V++ GL+GVWMAT+ DN RW L H+ +Y
35  Sbjct: 367  GKAKLPHYATTIGMWVIRIGLVGVVWQYGLIGVWMATVLDNTSRWFILSKHFVKY 424
    Identities = 48/211 (22%), Positives = 89/211 (41%), Gaps = 29/211 (13%)

    Query: 213  MTWKISAQIWNALPSAGERLMMRAGDVVIVAIVVQLGTNVVAGNAIGETLTQFNYPGL 272
    M + +I++LALPS E ++ +V +V Q+G V+G +I + +
40  Sbjct: 1  MIYNNRRKIFSLALPSMIENILQMLMGVVDNYLVAQIGLVAVSGVSIANNIISIYQSLFI 60

    Query: 273  GIATATIILTAKYVGQKNRESIEETIQSSYYIGLVLMILISSFML-----L 318
    + A L A+ +G+ N+ Q +Y G++ + L+ S L L
45  Sbjct: 61  ALGAAVSSLIARSIGENNQNK-----QLNYMAGVLQVTLTLLSVGLGLLSVAGHHQVLEWL 115

    Query: 319  AGKPLTQLFTNNPSAIKGLSLIVILLSFVGVDPATIGTLVYTAAWQGLGNAKLPHYTTTIGM 378
    + L +I G +IV L G+ ++G +V + G K+P + + +
50  Sbjct: 116  GAEASVTLVGGQYLSIVGGMIVSL----GLTSLGAIV----RAQGYPKIPMQVSLI-I 165

    Query: 379  WLIRVVLGYLLGIVFELGGLGVWMATIADNI 409
    ++ + L V+ GLLGV AT+ +
55  Sbjct: 166  NVLNAIFSALSIYVWGFGLLGVAWATVLSRL 196

```

A related GBS gene (SEQ ID 8071) and protein (SEQ ID 8072) were also identified. Analysis of this

-2339-

Example 2073

A DNA sequence (GBSx2188) was identified in *S.agalactiae* <SEQ ID 6417> which encodes the amino acid sequence <SEQ ID 6418>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2200 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD05671 GB: AE001448 THREONINE SYNTHASE [Helicobacter pylori
 J99]
 Identities = 161/479 (33%), Positives = 259/479 (53%), Gaps = 17/479 (3%)

Query: 14 KVTASQAILKGLADDGGLFTPTTFPKVDLDFTKLKDASYQEVAKLVLSAFFDDFTEQELD 73
 K+ +A+L A GGL+T F L++ SY E+ + V + + L
 Sbjct: 13 KIDFIEAVLNPNAPKGGLYTLEHFET--LEWQDCLGMSYSELVEHVFEELNLEIPKNLLA 70

Query: 74 YCISQAYDTKFDTTEIAPIVKIGDRYHL-ELFHGPTIAFKDMALSILPYLLTTAAKKQGV 132
 + + Y+ + API + +R + EL+HGP++AFKDMAL L L + A G
 Sbjct: 71 SALKR-YENFDNPKNPAPIFALNERLFVQELYHGPSLAFKDMALQPLASLFSNLAV--GK 127

Query: 133 DNKIVILTATSGDTGKAAMAGFADVPGTEIIVFYPKNGVSYIQELQMITQAGQNTHVVAI 192
 + K ++L +TSGDTG A + G A +P ++ YPK+G S +Q+LQM+TQ N V +
 Sbjct: 128 NEKYLVLVSTSGDTGPATLEGLAGMPNVFVVCYLPKDGTSLVQKLQMVITQNASNLKVFGV 187

Query: 193 EGNFDDAQTSVKEMFNNSLLRLKLSQQHMQLSSANSNMNIGRLVPQIVYIYIYAAQLVKSK 252
 G+FDDAQ ++K + + L + ++LS ANS+N GR+ QIVY+I+ + +L K
 Sbjct: 188 SGDFDDAQNALKNLLKDDDFNEALKARQLKLSVANSVNFGRIFQIVYHIWGFELEYKKG 247

Query: 253 EISIGQPINFSVPTGNFNGNLAAYASQIGLPVTKLICASNDNNVLTDFFKTQTYD-KNR 311
 I+ + I ++P+GNFGN L A+YA ++GL + K+ +N N+VL +F +T YD R
 Sbjct: 248 AINSKEKITLAIPSGNFGNALGAFYAKKMGLNIAKIKVVTNSNDVLRFEITETGRYDLTKR 307

Query: 312 EFKVTSSPSMDILVSSNLERLIFHLLGDDAETTKKLMDLVTTGGEYALEARQANIL-ESF 370
 K T SP+MDIL SSN+ER +F L G E T +LM+ L YAL+ ++ +L E F
 Sbjct: 308 SLKQTFSPAMDILKSSNVERALFSLFG--FERTLELMQALBEEKFYALKPKELALLQEHF 365

Query: 371 VAGFATEQFVELDIKHLFDQYQYIEDPHTAVASAVYQAYQTETKDQTPAVIVSTASPYKF 430
 +++ I+ ++ ++QY+ DPHTA A K ++ +TAS KF
 Sbjct: 366 SCASCSDEDCCLKTIQEVYAEHQVLIDPHTAT-----ALNASLKTHEKTLVSATASYEKF 419

Query: 431 PCVVTKAIT-NKEEIQDFAAISILNDLSGVSLPKAVIDLQAEVIHRTVVPTSNNMRETV 488
 P A+ K+ D AA+ L + + + DL + + H+ V+ + ++ ++
 Sbjct: 420 PKTTLALNEQKKNDDKALETLNKSYNTPDSQRLDDLFERGIKHQEVVLKLNKSI 478

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2074

A DNA sequence (GBSx2189) was identified in *S.agalactiae* <SEQ ID 6419> which encodes the amino acid sequence <SEQ ID 6420>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2340-

bacterial cytoplasm --- Certainty=0.3153(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9279> which encodes amino acid sequence <SEQ ID 9280> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF40975 GB:AE002410 alcohol dehydrogenase, propanol-preferring
 [Neisseria meningitidis MC58]
 10 Identities = 202/282 (71%), Positives = 228/282 (80%), Gaps = 1/282 (0%)

Query: 1 MGHEGIGIVEEIGEGVTSRLVGDVRSIAWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 60
 +GHEGIG+V+E+ +GV +L+VGDRVSIW F+ CG CEYC TGRETLCRSV NAGY+ DG
 15 Sbjct: 60 LGHEGIGLVKEVADGVKNLKVGDVRSIAWLFQSCGSCYECNTGRETLCRSVLNAGYTADG 119

Query: 61 GMSEYAIVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIAVYGAGGLGN 120
 GM+ + IV+ADYAVKVPEGLDPAQASSITCAGVTTYKAIK +G PGQWIA+YGAGGLGN
 15 Sbjct: 120 GMATHCIVSADYAVKVPEGLDPAQASSITCAGVTTYKAIKVSGVRPGQWIAIYGAGGLGN 179

Query: 121 LAVQYAKKVFNAHVVAVDINADKLQLAKEVGADLTVNGKEIKDVAAYIQEKTGGCHGVVV 180
 L VQYAKKVF AHVVA+DIN DKL AKE GADL VN + +D A IQEKTGG H VV
 20 Sbjct: 180 LGVQYAKKVFGAHVVAIDINDDKLAFKETGADLVVNAAK-EDAAKVIQEKTGGAAHAVV 238

Query: 181 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIKTVLDGIRVVGSLVGTTRKDLLEAF 240
 TAVS AFN A++ VRAGG VVA+GLP E M+LSI + VLDGI VVGSLVGTTRKDLLEAF
 25 Sbjct: 239 TAVSAAAFNSAVNCVRAGGRVVAIGLPPESMDLSIPRLVLDGIEVVGSLVGTTRKDLLEAF 298

Query: 241 AFGAEGLVVPVVEKVPVDTAPQVFDEMERGLIQGRKVLDFTK 282
 FGAEGLVVP V+ +D AP +F EM G I GR V+D K
 30 Sbjct: 299 QFGAEGLVVPVKVQLRALDEAPAIFQEMREGKITGRMVIDMKK 340

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6421> which encodes the amino acid sequence <SEQ ID 6422>. Analysis of this protein sequence reveals the following:

Possible site: 31
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2356(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 263/280 (93%), Positives = 273/280 (96%)

45 Query: 1 MGHEGIGIVEEIGEGVTSRLVGDVRSIAWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 60
 +GHEGIGIVEEIGEGVTSRL+VGDRVSIWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG
 Sbjct: 76 LGHEGIGIVEEIGEGVTSRLKVGDVRSIAWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 135

Query: 61 GMSEYAIVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIAVYGAGGLGN 120
 GMSEYA+VTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWI ++GAGGLGN
 50 Sbjct: 136 GMSEYAVVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIVIFGAGGLGN 195

Query: 121 LAVQYAKKVFNAHVVAVDINADKLQLAKEVGADLTVNGKEIKDVAAYIQEKTGGCHGVVV 180
 LAVQYAKKVFNAHVVAVDIN DKL+LAKEVGAD+ VNGKEI+DV YIQEKTGG HGVVV
 55 Sbjct: 196 LAVQYAKKVFNAHVVAVDINNDKLELAKEVGADILVNGKEIEDVPGYIQEKTGGAHGVVV 255

Query: 181 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIKTVLDGIRVVGSLVGTTRKDLLEAF 240
 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIKTVLDGI+VVGSLVGTTRKDLLEAF
 60 Sbjct: 256 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIKTVLDGIKVVGSLVGTTRKDLLEAF 315

Query: 241 AFGAEGLVVPVVEKVPVDTAPQVFDEMERGLIQGRKVLDF 280
 AFGAEGLV VPVVEKVPVDTAP+VFDEMERGLIQGRKVLDF

-2341-

Sbjct: 316 AFGAEGLVAPVVEKVPVDTAPEVFDemerGLIQGRKVLDF 355

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2075

A DNA sequence (GBSx2190) was identified in *Sagalactiae* <SEQ ID 6423> which encodes the amino acid sequence <SEQ ID 6424>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have a cleavable N-term signal seq.

10	INTEGRAL	Likelihood = -9.82	Transmembrane	83 - 99 (76 - 108)
	INTEGRAL	Likelihood = -7.27	Transmembrane	46 - 62 (43 - 65)
	INTEGRAL	Likelihood = -7.22	Transmembrane	187 - 203 (182 - 209)
	INTEGRAL	Likelihood = -6.00	Transmembrane	243 - 259 (229 - 262)
	INTEGRAL	Likelihood = -4.25	Transmembrane	404 - 420 (402 - 422)
15	INTEGRAL	Likelihood = -3.98	Transmembrane	120 - 136 (119 - 136)
	INTEGRAL	Likelihood = -3.88	Transmembrane	308 - 324 (307 - 324)
	INTEGRAL	Likelihood = -2.13	Transmembrane	378 - 394 (376 - 394)
	INTEGRAL	Likelihood = -1.38	Transmembrane	152 - 168 (152 - 168)
20	INTEGRAL	Likelihood = -1.17	Transmembrane	271 - 287 (271 - 287)

----- Final Results -----

	bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9371> which encodes amino acid sequence <SEQ ID 9372> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC17857 GB:AF026147 YojI [Bacillus subtilis]
 Identities = 183/432 (42%), Positives = 266/432 (61%), Gaps = 1/432 (0%)

Query: 1 MKLFIPVLIYQFANFSATFIDSVMTGQYSQHLHLAGVSTASNLWTPFFALLVGMISALVPV 60
 + + IP+ I Q TF+D+VM+G+ S LAGV+ S+LWTP + L G++ A+ P+

35 Sbjct: 15 LHILIPFITQAGLSLITFLDITVMSGKVSPADLAGVAIGSSLWTPVYTGLAGILMAVITPI 74

Query: 61 VGQHLGRGNKEQIRTEFHQFLYLGLILSLILFLIMQFIAQPVLGSLGLEDEVLA VGRGYL 120
 V Q LG K++I Q +Y+ +LS+ + +I +LG L L+ V + + +L

40 Sbjct: 75 VAQLLGAEEKKQKIPFTVLQAVYVAALLSIAVLVIGYAAVDLILGRNLNDIHVHQIAKHFL 134

Query: 121 NYMLIGIMPLVLFISICRSFFDALGLTRL SMLMLLILPFNSFFNYMLIYGKFGMPRLGGA 180
 ++ +GI PL ++++ RSF D+LG TR++M + L LP N NY+ I+GKFGMP LGG

45 Sbjct: 135 GFSLGIFLFLFVYTVLRSFIDSLGKTRVTMTLSSLPINFLNYVFIFGKFGMPALGGV 194

Query: 181 GAGLGTSITYWAFIVIIIVMSLHPQIKTYHIW-TLERIKAPLIIEDIRLGLPIGLQIFA 239
 GAGL ++LTYW I I+ ++ + Y I+ T+ + +++GLPIG +F

50 Sbjct: 195 GAGLASALTYWCICIIISFFIIHKNAFFSEYGIFLTMKFSWKACKNLLKIGLPIGFAVFF 254

Query: 240 EVAIFAVVGLFMAKFSSIIIAAHQAAMNFSSLMYAFPLSISTALAITISFEVGAERFQDA 299
 E +IFA V L M+ F ++ IA+HQAAMNF+SL+Y PLS+S AL I + FE GA RF+DA

55 Sbjct: 255 ETSIFAAVTLLMSHFHTVTIASHQAAMNFASLLYMLPLSVSMALTIVVGFEAGAARFKDA 314

Query: 300 NTYSRIGRLTAVGITSGLTLLFLFLFRENVAAMNSDPHFVAITAQFLTYSLFFQFADAYA 359
 +YS IG + A+G + T + LFRE +A MY SDP + +T FL Y+LFFQ +DA A

60 Sbjct: 315 RSYSLIGIMMAIGFSLFTAACILLFREQIAGMYTSDPDVLRITQHFLIYALFFQLSDAVA 374

Query: 360 APVQGILRGYKDTTKPFMIGAGSYWL CALPLAVILEKNSQLGPFAYWIGLITGIFVCGLF 419
 AP+QG LRGYKD SYW+ LP+ ++ + LG F YWIGLI G+ +

Sbjct: 375 APIQGALRGYKDVNYTLAAAFVSYWVIGLPVGYMVGTFSTLGAFGYWIGLIAGLAAGAVG 434

Query: 420 LNQLRLQKIKKLY 431

-2342-

L RL K++K Y
 Sbjct: 435 LFFRLAKLQKRY 446

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 5 vaccines or diagnostics.

Example 2076

A DNA sequence (GBSx2191) was identified in *S.agalactiae* <SEQ ID 6425> which encodes the amino acid sequence <SEQ ID 6426>. Analysis of this protein sequence reveals the following:

Possible site: 20
 10 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 23 - 39 (23 - 39)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2077

A DNA sequence (GBSx2192) was identified in *S.agalactiae* <SEQ ID 6427> which encodes the amino acid sequence <SEQ ID 6428>. Analysis of this protein sequence reveals the following:

Possible site: 52
 25 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.3829(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC06891 GB:AE000703 hypothetical protein [Aquifex aeolicus]
 35 Identities = 72/213 (33%), Positives = 115/213 (53%), Gaps = 11/213 (5%)
 Query: 36 RPKILMHVCCAPCSTYTLEYLSQ---WADVITYFANSNIHPKDEYYRREYVTQKFVHDFN 92
 + KIL+H+CCAP + Y L+ L + +++ YF + NIHP +EY R T++ +
 Sbjct: 3 KSKILVHICAPDAIYFLKKLREDYPESEIIGYFYDENIHPYEEYRLRYLETERICKELG 62
 40 Query: 93 KNTGYSVQFLSAPYEPNEFFKIVHGLEEEPEGGDRCKVCYDFRLDKTAEKAVELGFDYFG 152
 N + Y+ + + V G E+EPE G RC++C+D+RL+K+AE A ELG D
 Sbjct: 63 IN-----LIEGEYDLENWLERVKGYEDEPERGKRCQICFDYRLEKSAEVAKELGCDALT 116
 45 Query: 153 SALTISPHKNSQTINTIGIDVQKIYDTQYLPDLKKNKGYSVEMCKDYDIYRQCYCGC 212
 + L +SP K+ + G + K ++L D +K G Q ++ K+ +Y+Q YCGC
 Sbjct: 117 TTLMSPPKKSIPQLKKAGEEATKRTGIEFLAPDYRKGGGTQEMFKLSKEREIYQQDYCGC 176
 Query: 213 IFGAKDQGINLLQIKKDAKAFVSDKDGKEEFPN 245
 50 I+G Q +I D F+ + G +E N
 Sbjct: 177 IYGLFKQKNG--KIFWDLVGFLGRRPGSKEERN 207

-2343-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6429> which encodes the amino acid sequence <SEQ ID 6430>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3498(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 254-256

The protein has homology with the following sequences in the databases:

>GP:AAC06891 GB:AE000703 hypothetical protein [Aquifex aeolicus]
 Identities = 65/182 (35%), Positives = 106/182 (57%), Gaps = 9/182 (4%)

Query: 39 RPSILMHVCCAPCSTYTLEYLTQF---ADITVYFANSNIHPKDEYHRRAYVTQQFVSEFN 95
 + IL+H+CCAP + Y L+ L + ++I YF + NIHP +EY R T++ E
 Sbjct: 3 KSKILVHICCAPDAIYFLKKLREDYPESEIIGYFYDPNIHPYEEYRLRYLETERICKELG 62

Query: 96 AKTGNTVQFLEADYVPNEYVRQVRGLEEEPEGGDRVCFCFYRLDKTAQKAVELGFDYFA 155
 + +E +Y ++ +V+G E+EPE G RC++CFDYRL+K+A+ A ELG D
 Sbjct: 63 -----INLIEGEYDLENWLERVKGYEDEPERGKRCQICFDYRLKSAEVAKELGCCALT 116

Query: 156 SALTISPHKNSQTINDVGIDVQKVYTTKYLPSDFKKNNGYRRSVEMCEEYDIYRQCYCGC 215
 + L +SP K+ + G + K ++L D++K G + ++ +E +IY+Q YCGC
 Sbjct: 117 TTLMSPPKKSIPQLKKAGEEATKRTGIEFLAPDYRKGGGTQEMFKLSKERIYQQDYCGC 176

Query: 216 VY 217
 +Y
 Sbjct: 177 IY 178

An alignment of the GAS and GBS proteins is shown below.

Identities = 184/255 (72%), Positives = 219/255 (85%)

Query: 1 MIDVENILEKMKPNQKINYDWMQMQVKQWQASDIRPKILMHVCCAPCSTYTLEYLSQWA 60
 MID++ IL M PNQKINYD VMQMQ K W+ +RP ILMHVCCAPCSTYTLEYL+Q+A
 Sbjct: 4 MIDLQEILANMNPQKINYDRVMQMAKVWEKESVRPSILMHVCCAPCSTYTLEYLTQFA 63

Query: 61 DVTIYFANSNIHPKDEYRREYVTQKFVHDFNKNTGYSVQFLSAPYEPNEFFKIVHGLEE 120
 D+T+YFANSNIHPKDEY+RR YVTQ+FV +FN TG +VQFL A Y PNE+ + V GLEE
 Sbjct: 64 DITVYFANSNIHPKDEYHRRAYVTQQFVSEFNAKTGNTVQFLEADYVPNEYVRQVRGLEE 123

Query: 121 EPEGGDRCKVCYDFRLDKTAEKAVELGFDYFGSALTISPHKNSQTINTIGIDVQKIYDTQ 180
 EPEGGDRC+VC+D+RLDKIA+KAVELGFDYF SALTISPHKNSQTIN +GIDVQK+Y T+
 Sbjct: 124 EPEGGDRCRVCFCFYRLDKTAQKAVELGFDYFASALTISPHKNSQTINDVGIDVQKVYTTK 183

Query: 181 YLPSDLKKNKGYSQSVEMCKDYDIYRQCYCGCIFGAKDQGINLLQIKKDAKAFVSDKD GK 240
 YLPSD KKN GY+RSVEMC++YDIYRQCYCGC++ AK QGI+L+Q+KKDAKAF++DKD
 Sbjct: 184 YLPSDFKKNNGYRRSVEMCEEYDIYRQCYCGCVYAAKMQGIDLQVKKDAKAFMADKDL D 243

Query: 241 EEFPNIRFTFNGKSM 255
 +F +IRF++ G M
 Sbjct: 244 NDFTHIRFSYRGDEM 258

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2078

A DNA sequence (GBSx2193) was identified in *S.agalactiae* <SEQ ID 6431> which encodes the amino acid sequence <SEQ ID 6432>. Analysis of this protein sequence reveals the following:

-2344-

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4216(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB14809 GB:Z99118 excinuclease ABC (subunit C) [Bacillus subtilis]
 Identities = 189/333 (56%), Positives = 244/333 (72%)

Query: 1 MNELIKHKLELLPDSGPCYLHKDKNGTIIYVGKAKNLKRVKSYFHGSHNTKTELLVSEI 60
 MN+ +K KL LLED PGCYL KD+ T+IYVGKAK LKNRV+SYF GSH+ KT+ LV+EI

15 Sbjct: 1 MNKQLKEKLALLPDQPGCYLMDKDRQQTVIYVGKAKVLKRVRSYFTGSHDAKTQRLVTEI 60

Query: 61 EDFEYIVTTSNTEALLLEINLIQENMPKYNIRLKDDKSYPYIKITNERYPRLMITRQVK 120
 EDFEYIVT+SN EAL+LE+NLI+++ PKYN+ LKDDK+YP+IK+T+ER+PRL++TR VKK

20 Sbjct: 61 EDFEYIVTSSNLEALILEMNLIKKHDPKYNVMLKDDKTYPFIKLTHERHPRLIVTRNVK 120

Query: 121 SDGTIFYGPPYDPSGAATEIKRLLDRLFPFKKCTNPANKVCFYHLCNAHTVCQTNKAYW 180
 G YFGYPY+ AA E K+LLDRL+P +KC+ ++VC YYHLGQC A V ++

25 Sbjct: 121 DKGRYFGPPYVQAAARETKKLLDRLYPLRKCSKLPDRVCLYHLCQCLAPCVKDISSEETN 180

Query: 181 DSLREDVQQLNGKDNKIVNGLTEKMKSAAAMTEFFERAAEYRDLIEAISLLRTQQRVIHQ 240
 L E + +FL G N++ L EKM AA +EFERA E RD I I KQ++

30 Sbjct: 181 RELVESITRFLRGGYNEVKKELEEKMEAAENLEFERAKELRDQIAHIESTMEKQKMTMN 240

Query: 241 DMKDRDVFYGFYVDKGWMCVQVFFVNGKLIQRDVMFPYYNEPEEDFLTYIGQFYQDTKH 300
 D+ DRDVF Y DKGWMCVQVFF+R GKLI+RDV+MFP Y E +E+FLT+IGQFY H

35 Sbjct: 241 DLVDRDVFAYAYDKGWMCVQVFFIRQGLIERDVSMPFLYQEADEEFLTFIGQFYSKNNH 300

Query: 301 FLPKVEFIPQDIDAKSVETIVGCKIVKPQRGKR 333
 FLPKE+ +P ID +E ++ + +P++G +

35 Sbjct: 301 FLPKEILVPDSIDQSMIEQLLETNVHQPCKGPK 333

There is also homology to SEQ ID 2568.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2079

A DNA sequence (GBSx2194) was identified in *Sagalactiae* <SEQ ID 6433> which encodes the amino acid sequence <SEQ ID 6434>. This protein is predicted to be maltose operon transcriptional repressor (rbsR). Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3761(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9393> which encodes amino acid sequence <SEQ ID 9394> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAD02112 GB:AF039082 putative maltose operon transcriptional
 repressor [Lactococcus lactis]
 Identities = 64/166 (38%), Positives = 105/166 (62%), Gaps = 13/166 (7%)

-2345-

Query: 1 MGKSAIDYLYKKGHKSIQFVTDDLNSEVSEERYLGYPFGARKLGLNQKPALLFDRGNPQV 60
 +G+ A+ L + H++I FVTD +EV EERY G+ A +LGL+ LLF N +
 Sbjct: 169 LGREAVRLLAQLNHQNISFVTDTKETEVEERYQGFKDEAERLGLSHD--LLFMDSNFSL 226

Query: 61 LEEFINRVKEEETLALIVIGDITVSVRVMQFLSFYKLVKVPDDISIMTFNNSLFSHLIHPYL 120
 E TAL+V+ D +S++V++ L L VP+D+S++T+NNS+F +IHPYL
 Sbjct: 227 RNE-----TALVVMDDVLSLKVVERLRSQGLNVPEDVSLITYNNSIFGAMIHPYL 276

Query: 121 STFDINVNNGRTSVRRLLIDIKSPDKVFSETIIVPFTLEERESVR 166
 +TFDI++ LG ++++++D+ + + + +TII PF L RES +
 Sbjct: 277 TTFDIHIEQLGASAIKILDRLDNKENLPEKTII-PFELIVRESTK 321

There is also homology to SEQ ID 5082.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2080

A DNA sequence (GBSx2195) was identified in *S. agalactiae* <SEQ ID 6435> which encodes the amino acid sequence <SEQ ID 6436>. This protein is predicted to be 4-alpha-glucanotransferase (malQ). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2003(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26923 GB:J01796 amylomaltase [Streptococcus pneumoniae]
 Identities = 250/500 (50%), Positives = 329/500 (65%), Gaps = 4/500 (0%)

Query: 1 MKKRASGVLHMTSLPGDLGIGTFGREAYAFVDFLVETDQKFWQILPLITTSFGDSPYQS 60
 MKKR SGVLMHI+SLPG GIG+FG+ AY FVDFLV T Q++WQILPL TS+GDSPLYQS
 Sbjct: 1 MKKRQSGVLMHISSLPGAYGIGSGFGQSAFYDFVDFLVRTKQRYWQILPLGATSYGDSPLYQS 60

Query: 61 FSAVAGNTHLIDFDLLTLEGFISKDDYQNIISFGQDPEVVDYAGLFKRRPVLKAVKNFL 120
 FSA AGNTH ID D+L +G + D + + FG D VDYA ++ RRP+LEKAVK F
 Sbjct: 61 FSAFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFF 120

Query: 121 QEERATRLMSDFLQE-EKWTDFAEFMAIKEHFGNKALQEWDDKAIIRBEEALAGYRQK 179
 E + F Q+ + W+ FAE+MAIKE+F N A EW D R+ AL YR++
 Sbjct: 121 -EVGDVKDFEKFQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALESYREQ 179

Query: 180 LSEVIKYHEVTQYFFYKQWFELKEYANDKGIQIIGDMPIYVSADSVEVWMPPELFKLD RD 239
 L++ + YH VTQYFF++QW +LK YAND I+I+GDMPIYV+ DS ++W P LFK D +
 Sbjct: 180 LADKLVYHRVTQYFFQQLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN 239

Query: 240 KQPLAIAAGVPADDFSDGQLWGNPIYNWDYHKESDFDWWIYRIQSGVKMYDYLRIDHFKG 299
 + IAG P D+FS GQLWGNPIY+W+ + + WWI R++ K+YD +RIDHF+G
 Sbjct: 240 GKATCIAGCPPDEFBSVTGQLWGNPIYDWEAMDKGYKWWIERLRESFKIYDIVRIDHFRG 299

Query: 300 FSDYWEIRGIDYQTANDGSWQAPGPELFATIKEKLGDLPIIAENLGYIDERAERLLAGTG 359
 F YWEI TA G W PG +LFA +KE+LG+L IIAE+LG++ + L TG
 Sbjct: 300 FESYWEIPAGSDTAAPGEWVGPGYKLFAAVKEELGELNIIAEDLGFMTEVIELRERTG 359

Query: 360 FPGMKIMEFGFYDTTGNISIDIPHNYTENTIAIYAGTHDNEVINGWFEN-LTVEQKAYAENY 418
 FPGMKI++F F + SID PH N++ Y GTHDN + GW+ N + + Y Y
 Sbjct: 360 FPGMKILQFAF-NPEDESIDSPLAPANSVMYTGTHDNNTVLGWYRNEIDDATREYMARY 418

-2346-

Query: 419 MRRLPNEPITETVLRTLYATVSQTTITCMQDLLDKPADSRMNPNTVGGNWQWRMRKEDL 478
 R E + +LRT+++VS I MQDLL+ +RMN P+T+GGNW WRM ++ L
 Sbjct: 419 TNRKEYETVVHMLRTVFSSVSFMAIATMQDLELDEAARMNFPSTLGGNWSWRMTEDQL 478

5 Query: 479 TENRKAFLKEITTIYNRGNK 498
 T + L ++TTIY R N+
 Sbjct: 479 TPAVEEGLLDLTTIYRRINE 498

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6437> which encodes the amino acid
 10 sequence <SEQ ID 6438>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.85 Transmembrane 435 - 451 (435 - 451)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 313/495 (63%), Positives = 387/495 (77%)

Query: 1 MKKRASGVLMHITSLPGDLGIGTFGREAYAFVDFLVETDQKFWQILPLTTTSFGDSPYQS 60
 M KRASG+LMHI+SLPG GIGTFG+ A+ FVDFL ET Q +WQILPLTTTSFGDSPYQS
 25 Sbjct: 1 MNKRASGILMHISLPGKFGIGTFGKSAFEVDFLAETKQTYWQILPLTTTSFGDSPYQS 60

Query: 61 FSAVAGNTHLIDFDLLTLEGFISKDDYQNISFGQDPEVVDYAGLFEKRRPVLEKAVKNFL 120
 FSA+AGNTH ID+LL + + D +I+FG +PE VDYA LF+ RRP+LEKAV+ F+
 30 Sbjct: 61 FSAIAGNTHFIDFELLVDELLEAADLCDITFGTNPEAVDYAQLFQVRRPLLEKAVRAFV 120

Query: 121 QEERATRLMSDFLOEEKWVTDFAEFMAIKEHFGNKALQEWDKAIIRREEEALAGYRQKL 180
 E+ L F W+TDAEFMA+KE+F NKALQ+WDD+ +I+R+E++L YR+ L
 Sbjct: 121 AEQENVCKLEAFETASSWLTDFAEFMALKEYFNKALQDWDDETVIKQBEDSLNNYRELL 180

35 Query: 181 SEVIKYHEVTQYFFYKQWFELKEYANDKGIQIIGDMPYVSADSVEVWMPPELFLDRDK 240
 ++ I YH+V QYFFY+QW LK YAN KGI+IIGDMPYVSADSVEVWMPPELFLK+D DK
 Sbjct: 181 AKKITYHKVCQYFFYQWSALKTYANHKGIEIIGDMPYVSADSVEVWMPPELFLKVDSDK 240

40 Query: 241 QPLAIAGVPADDFSDGQLWGNFIYNWDYHKESDFDWWIYRIQSGVKMYDYLRIDHFKGF 300
 +FL IAGVPAD FS+DGQLWGNP YNW H++S+F WWIYRIQ K+YD LRIDHFKGF
 Sbjct: 241 KPLFIAGVPADGFSEDGQLWGNPTYNWSAHEKSNFAWWIYRIQESFKLYDQLRIDHFKGF 300

Query: 301 SDYWEIRGDYQTANDGSWQAPGPPELFATIKEKLGDLPIIAENLGYIDERAERLLAGTGF 360
 SD+WEI +TA +G W APG LF+ ++E LG+LPIIAENLGYIDE+AE+LLA TGF
 45 Sbjct: 301 SDFWEIPAGDKTARNGHWASAPGIALFSAVREALGELPIIAENLGYIDEKAEQLLASTGF 360

Query: 361 PGMKIMEFGFYDITGNSIDIPHYNTENTIAYAGTHDNEVINGWFENLTVEQKAYAENYMR 420
 PGMKI+EFG +D T SID+PH Y N +AY GTHDNEV+NGW++NL+ EQ + NY+
 50 Sbjct: 361 PGMKILEFGLFDITSQSIDLPHYDRNCVAYTGTHDNEVNGWYDNLSEEQVHFVNNYLH 420

Query: 421 RLPNEPITETVLRTLYATVSQTTITCMQDLLDKPADSRMNPNTVGGNWQWRMRKEDLTE 480
 + +E IT+ +LRT++A+V T I C+QDLLDK SRMNPNT+GGNWQWRM +L +
 Sbjct: 421 KHADESITKAMLRITIFASVCDTAILCTQDLLDKDGKSRMNPNTIGGNWQWRMLDGLNLQ 480

55 Query: 481 NRKAFLKEITTIYNR 495
 + K +L +T +Y R
 Sbjct: 481 DHKDYLIYLTLDLYGR 495

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 60 vaccines or diagnostics.

-2347-

Example 2081

A DNA sequence (GBSx2196) was identified in *S.agalactiae* <SEQ ID 6439> which encodes the amino acid sequence <SEQ ID 6440>. This protein is predicted to be glycogen phosphorylase (malP). Analysis of this protein sequence reveals the following:

```

5   Possible site: 40
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2678(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAC00218 GB:AF008220 glycogen phosphorylase [Bacillus subtilis]
    Identities = 297/776 (38%), Positives = 452/776 (57%), Gaps = 41/776 (5%)

    Query: 13  GKVLSELTNEEIYVELLN FVKEEAAA-----KSKNSSQRKVYYISAEFLIGKLLSNNL 65
              GK  +   + Y  L N V+E  +A      KS+++S ++ YY+S EFL+G+LL  NL
    Sbjct: 21  GKSFKDSAKLDQYKTLGNMVREYISADWIETNEKRSRSNSGKQTYLSTIEFLGQLEQLN 80

20   Query: 66  INLGIYKDVKKELELVGKSIABIEDVEPEPSLNGGGLGRLASC FIDSISSLGINGEGVGL 125
              +NLG+  V+  L+ +G ++ EI  +E +  LGNGGLGRLA+CF+DS++SL + G G+G+
    Sbjct: 81  MNLGVRDVVEAGLKEIGINLEELQIENDAGLNGGGLGRLAACFLDSLASLNLPGHGMGI 140

25   Query: 126 NYHCGLFKQVFRNNQQEAEANYWIEN-NSWL VPT-DISYDVPF-----RDFTLKSR 175
              Y  GLF+Q  + Q      W++N N W V  D + DVPF      +  L  R
    Sbjct: 141 RYKHGLFEQKIVDGHQVELPEQWLKNGNVWEVRNADQAVDVPFWEVHMTEKSGRLHFRH 200

30   Query: 176 DR-----IDVLGYKKDTKNYLNLFIDGLDYNLIEKGITFDKTEIKKNLTLFLYP 225
              ++      I ++GY+  T N L L++ +  Y  G      + ++ FLYP
    Sbjct: 201 EQATIVTAVPYDIFIIGYETGTVNTRLRLWNAE--PYAHYHGGNLSYKRETEAVSEFLYP 258

    Query: 226 DSDKNGELLRIYQQYFMVSNAAQLLIDEAIERGSNLDLA EYAVQINDTHPSMVIPEL 285
              DD+  G++LR+ QQYF+V  + + +++  +  +L  L +  + INDTHP++ +PEL
35   Sbjct: 259 DDTHEGKILRLKQQYFLVCASLKSIVNNYRKTHKSLSGLHKKVS IHINDTHPALAVPEL 318

    Query: 286 IRLLTEKHGFEFDEAVSVVRNMVGYTNHTILAEALEKWPLEYLNEVPHLVTI IKLDQM 345
              +R+L ++      ++EA  +  + + YTNHT L+EALEKWP+      ++P +  II+++++
40   Sbjct: 319 MRILDEENMSWEEAWHITVHTISYTNHTLSEALEKWPIHLFKPLLRMYMIIIEINER 378

    Query: 346 IRE-----EQTNPEVQIIDEAGRVHMAHMDIH FSTSVNGVAALHTEILKNSLKV FY 397
              +      E  I  G V MAH+ I  S SVNGVA +H++ILK  E++ F+
    Sbjct: 379 FCRAVWEKYPGDWKRIENMAITAHGVVKMAHLAIVGSYSVNGVAKIHS DILKEREMRDFH 438

45   Query: 398 DIYPDKFNKNTNGITFRRWLEFANQDLADYLKELIGDSYLT DATQLEKLLTYADSNEVHD 457
              ++P+++FNKNTNGI  RRWL  AN  L+  + E IGD ++      L +L  YA      +
    Sbjct: 439 LLFPNRFNNTNGIAHRRWLLKANPGLSAIITEAIGDEWVKQPESLIRLEPYATDPATFE 498

50   Query: 458 KLAAIKFNKLALKRYLKENKGIELDEYSIIDTQIKRFHEYKRQOMNALYVIHKYLEIKR 517
              +  K K K  L  +      G+ ++  SI D Q+KR H YKRQ +N L++++ Y  +K
    Sbjct: 499 QFQNNKSKKKQELADLIFCTAGVVVN PESIFDVQVKRLHAYKROLLNLVHIMYLYNRLKE 558

    Query: 518 GH-FPSRKLTVIFGGKAAPAYTIAQDIH LILCLSELINNDPEVNKYLVNHLVENYNVTV 576
              F      T IFG KA+P+Y  A+ II LI  ++E +N DP V +  + V  +ENY V++
55   Sbjct: 559 DSGFSIYPQTFIFGAKASPSYYYAKIIKLIHSVAEKVNYDPAVKQLIKVVFL ENYRVSM 618

    Query: 577 AEKLIPATDISEQISLASKEASGTGNMKFMLNGALT LTGTMDGANVEIAELAGKENIYTFG 636
              AE++ PA+D+SEQIS ASKEASGTGNMKFM+NGALT+GT DGAN+EI E  G  + IYTFG
60   Sbjct: 619 AERIFPASDVSEQISTASKEASGTGNMKFMMNGALTIGTHDGANIEILERVGPDCIYTFG 678

    Query: 637 KSDSTIINLYETSGYRSKDYD KDKVIREAVDFIISDDIVSLGNAERLKR LHDELV-GKD 695
              +D +++  E  GYRS++YY  D+ IR+  D +I+      G A+  + + D L+  D
    Sbjct: 679 LKADEVLSYQENGGYRSREYYQHRRIRQVADQLINGFFE--GEADEFESIFDSLPHND 736

```

-2348-

Query: 696 WFM TLIDLKEYI AVKEQVLADYEDYESWNKKVIHNI AKAGFFSSDRTIEQYNQDIW 751
 + L D Y +E++ ADY + W++ I NIA +G+FSSDRTI +Y +DIW
 Sbjct: 737 EYFVLKDFSSYADAQERI QADYRERRKWEHSI VNI AHSGYFSSDRTIREYAKDIW 792

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6441> which encodes the amino acid sequence <SEQ ID 6442>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.71 Transmembrane 538 - 554 (538 - 554)

----- Final Results -----

bacterial membrane --- Certainty=0.2084(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 629/754 (83%), Positives = 696/754 (91%), Gaps = 2/754 (0%)

20 Query: 1 MTRNFTTYVGQQ-GKVLSELTNEEIIYVELLN FVKEEAAKSKNSSQRKVYVYISAEFLIGK 59
 MTR FT YV + GK L++ +NEEII+ LLNFVKEEA+ K+KNS++RKVYVYISAEFLIGK
 Sbjct: 1 MTR-FTEYVETKLGKSLTQASNEEIIYLSLLNFVKEEASHKAKNSAKRKVYVYISAEFLIGK 59

25 Query: 60 LLSNNLINLGIYKDVKKELELVGKSIAETEDVEPEPSLGN GGLGRLASCFIDSISSLGIN 119
 LLSNNLINLGIYKD+K+EL GKSIAE+EDVE EPSLGN GGLGRLASCFIDSI+SLGIN
 Sbjct: 60 LLSNNLINLGIYKDIKEELAAAGKSIAEVEDVELEPSLGN GGLGRLASCFIDSIASLGIN 119

30 Query: 120 GEGVGLNYHCGLFKQVFRNNQOEAEANYWIENNSWL VPTDISYDVFFRDFTLKSRLDRID 179
 GEGVGLNYHCGLFKQVF++N+QEAE N+WIE++SWLVPTDISYDVFF++FTLKSRLDRID
 Sbjct: 120 GEGVGLNYHCGLFKQVFKHNEQEAEPNFWIEDDSWL VPTDISYDVFFKNFTLKSRLDRID 179

35 Query: 180 VLGYKDKTKNYLNLFDIDGLDYNLIEKGITFDKTEIKKNLTLFLYPDDSDKNGELLRIYQ 239
 VLGYK+DTKKNYLNLFDI+G+DY LI+ GI+FDKT+I KNLTFLFLYPDDSDKNGELLRIYQ
 Sbjct: 180 VLGYKRDTKNYLNLFDIEGVGYGLIKDGISFDKTQIAKNLTLFLYPDDSDKNGELLRIYQ 239

40 Query: 240 QYFMVSNA AQLLIDEA IERGSNLHDLAEYAYVQINDTHPSMVIPELIRLLTEKHGFEFDE 299
 QYFMVSNA AQL+IDEA IERGSNLHDLA+YAYVQINDTHPSMVIPELIRLLTEKHGF+FDE
 Sbjct: 240 QYFMVSNA AQLIIDEA IERGSNLHDLADYAYVQINDTHPSMVIPELIRLLTEKHGFDFDE 299

45 Query: 300 AVSVVRNMVGYTNHTILAEALEKWPLEYLNEVVP HLVITIIKKLDQMIREEQTNPEVQIID 359
 AV+VV+NMVGYTNHTILAEALEKWP YLNEVVP HLVITII+KLD ++R E ++P VQIID
 Sbjct: 300 AVAVVKNMVGYTNHTILAEALEKWP TAYLNEVVP HLVITIEKLDALVRSEVSDPAVQIID 359

50 Query: 360 EAGR VHMAHMDIHFSTSVNGVAALHTEILKNSELKV FYDIYPDKFNKNTNGITFRRWLEF 419
 E+GRVHMAHMDIHF+TSVNGVAALHTEILKNSELK FYD+Y+P+KFNKNTNGITFRRWLEF
 Sbjct: 360 ESGRVHMAHMDIHFATSVNGVAALHTEILKNSELKAFYDLYPEKFNKNTNGITFRRWLEF 419

55 Query: 420 ANQDLADY LKELIGDSYLT DATQLEKLLTYADSNEVHDKLA AIKFKNKLAL KRYLKENKG 479
 ANQDLADY+KELIGD YLT DAT+LEKL+ +AD VH KLA IKF NKLAL KRYLK+NK
 Sbjct: 420 ANQDLADYIKELIGDEYLT DATKLEKLMAFADDKAVHAKLABIKFNKKLAL KRYLKDNDK 479

60 Query: 480 IELDEYSIIDTQIKRFHEYKRQ QMNALYVIHKYLEIKRGHFPSRKLTVIFGGKAAPAYII 539
 IELDE+SIIDTQIKRFHEYKRQ QMNALYVIHKYLEIK+G+ P RK+TVIFGGKAAPAY I
 Sbjct: 480 IELDEHSIIDTQIKRFHEYKRQ QMNALYVIHKYLEIKKGNLPRKKTITVIFGGKAAPAYII 539

65 Query: 540 AQDIIHLILCLSELINNDPEVNKYLN VHLVENYNVTVAEKLIPATDISEQISLASKEASG 599
 AQDIIHLILCLSELINNDPEV+ YLN VHLVENYNVTVAE LIPATDISEQISLASKEASG
 Sbjct: 540 AQDIIHLILCLSELINNDPEVSPYLN VHLVENYNVTVAEHLIPATDISEQISLASKEASG 599

Query: 600 TGNMKFMLNGALT LGTMDGANVEIAELAGKENIYTFGKSDTIINLYETSGYRSKDYYDK 659
 TGNMKFMLNGALT LGTMDGANVEIAELAG ENIYTFGKSDTIINLY T+ Y +KDYD
 Sbjct: 600 TGNMKFMLNGALT LGTMDGANVEIAELAGMENIYTFGKSDTIINLYATASYVAKDYDYN 659

Query: 660 DKVIREAVDFIISDDIVSLGNAERLKR LHDELVGKDWFM TLIDLKEYI AVKEQVLADYED 719
 I+ AV+FIIS ++++ GN ERL RL+ EL+ KDWFM TLIDL+BYI VKE++LADYED
 Sbjct: 660 HPAIKAAVNFIISP ELLAFGNEERLDRLYKELISKDWFM TLIDLEEYIEVKEKMLADYED 719

-2349-

Query: 720 YESWNKKVIHNIAGKAGFFSSDRTIEQYNQDIWHS 753
 + W KV+HNIAGKAGFFSSDRTIEQYN+DIWHS
 Sbjct: 720 QDLWMTKVHNIAGKAGFFSSDRTIEQYNEDIWHS 753

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2082

A DNA sequence (GBSx2197) was identified in *S. agalactiae* <SEQ ID 6443> which encodes the amino acid sequence <SEQ ID 6444>. This protein is predicted to be glycerol-3-phosphatase transporter (glpT). Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

15	INTEGRAL	Likelihood = -10.88	Transmembrane	339 - 355 (333 - 359)
	INTEGRAL	Likelihood = -8.01	Transmembrane	432 - 448 (426 - 450)
	INTEGRAL	Likelihood = -6.74	Transmembrane	92 - 108 (91 - 127)
	INTEGRAL	Likelihood = -6.69	Transmembrane	194 - 210 (190 - 214)
	INTEGRAL	Likelihood = -3.77	Transmembrane	367 - 383 (364 - 385)
	INTEGRAL	Likelihood = -2.81	Transmembrane	111 - 127 (109 - 127)
20	INTEGRAL	Likelihood = -2.28	Transmembrane	407 - 423 (406 - 424)
	INTEGRAL	Likelihood = -2.02	Transmembrane	165 - 181 (165 - 182)
	INTEGRAL	Likelihood = -0.64	Transmembrane	29 - 45 (29 - 45)

----- Final Results -----

25	bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAC44575 GB:U28354 IS629 ORFB fused with sequences similar to E.
 coli GlpT and UhpT proteins, Swiss-Prot Accession Number
 P08194 and P09836; Method: conceptual translation
 supplied by author [Shig
 Identities = 174/321 (54%), Positives = 241/321 (74%), Gaps = 4/321 (1%)

35 Query: 109 GVIPSVITSIWLFITIMYLINGWLQGMGYPPGARTLVVYWDNKERIKYATIWNLSHNFSGA 168
 GV P V + + + YL+NGW+QGMGYPPGA+TLV+WY+++ERI +AT+WNLSHN GGA
 Sbjct: 12 GVGP-VCSELHIAPSTYLLNGWIQGMGYPPGAKTLVFWYEHRRERISWATLWNLSHNVGGA 70

40 Query: 169 IAPILTGVLALAGNDSLNQARAAYWFGVVACLLAVLVYFLQEDTPESIGLPPIEEYHK 228
 +AP+L G G+ +L+ ARAA+ FPGV+ ++VL+YF+Q D P S+GLPPIEE+
 Sbjct: 71 LAPVLIGFSFGFFGDSALDHARAAFIFPGVLCMAMSVLIYFIQVDRPVSVGLPPIEEWKG 130

45 Query: 229 EQYTNVVDSSDILEEPEVLGMGEIHKYILPNTKLMWASLYSIFVYILRYGIVSWTPKFL 288
 ++ E+ L + +II+K+I+ N KL++ +Y FVYILRYGIVSW PKFL
 Sbjct: 131 NVVSHPAKGR---EQGPRLSIPDIIRKHIIRNNKLIYCCIYGSFVYILRYGIVSWAPKFL 187

50 Query: 289 ATSVQDGGKGITATAGMGGFSLFEIGGIIGMLTAGYLSAKVFKNKPLTNVAFVVAAIL 348
 + S+ GGK + A MGG S+FEIGG+ GML AGYLS ++F+NSKPLTN FL + I+L
 Sbjct: 188 SDSLDVGGKDMGLASMGGSVFIEIGGVAGMLLAGYLSVRLFRNSKPLTNTLFLALTIIL 247

55 Query: 349 LAAYWFIPAGPQYMALDFIILLGLGASIYGPVMMVGLYAMELVPKAAAGAASGLTGTFYSY 408
 L AYW++P+G +Y+ L++ IL+ LG ++YGPVM +GLY+MELVPK AAGAASGL+GTFYSY
 Sbjct: 248 LIAYWVPSGNEYLWLNYYTILILLGLAVYGPVMEFIGLYSMELVPKEAGAASGLSGTFYSY 307

Query: 409 VGGATIATLAIGIIDIHFGWG 429
 + G+ +ATL +G+++D+ GWG
 Sbjct: 308 IFGSIVATLGMGLVVDYLGWG 328

-2350-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6445> which encodes the amino acid sequence <SEQ ID 6446>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

5	INTEGRAL	Likelihood = -12.37	Transmembrane	185 - 201 (175 - 208)
	INTEGRAL	Likelihood = -9.13	Transmembrane	114 - 130 (90 - 134)
	INTEGRAL	Likelihood = -7.75	Transmembrane	322 - 338 (320 - 345)
	INTEGRAL	Likelihood = -6.79	Transmembrane	421 - 437 (419 - 439)
	INTEGRAL	Likelihood = -6.37	Transmembrane	91 - 107 (90 - 113)
10	INTEGRAL	Likelihood = -5.36	Transmembrane	163 - 179 (161 - 181)
	INTEGRAL	Likelihood = -5.20	Transmembrane	350 - 366 (347 - 371)
	INTEGRAL	Likelihood = -4.41	Transmembrane	23 - 39 (22 - 41)
	INTEGRAL	Likelihood = -3.77	Transmembrane	257 - 273 (249 - 273)
	INTEGRAL	Likelihood = -1.33	Transmembrane	61 - 77 (61 - 77)
15	INTEGRAL	Likelihood = -1.28	Transmembrane	383 - 399 (383 - 399)
	INTEGRAL	Likelihood = -0.90	Transmembrane	299 - 315 (299 - 315)

----- Final Results -----
 bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF96050 GB:AE004355 glycerol-3-phosphate transporter [Vibrio cholerae]
 Identities = 128/438 (29%), Positives = 215/438 (48%), Gaps = 17/438 (3%)

Query: 1 LFMEEDYNKREP-EKFTQFLRRQKVFFFAFF-GYVCAYLVRNNFKLMSNTIMVQNGWDK 58
 LF + +R P +K R + F+ F GY YL R NF L + +++ G+ +
 Sbjct: 21 LFKPAAHTQRLPSDKVDSVYSRLRWQLFIGIFVGYAGYYLGRKNFSL-AMPYLIEQGFSS 79

Query: 59 AQIAILLSCLTIVSYGLAKFYMGALGDRVSLRKLFSISLGASALICILIGFF---NSSMVV 115
 + + L ++++YGL+KF MG + DR + R S L SAL+ GF S+
 Sbjct: 80 GDLGVALGAVSIAYGLSKFLMGVSDRSNPRYFLSAGLLLSALVMFCFGFMPWATGSITA 139

Query: 116 LGILLVLCGVVQALAPASQAMIANYPNKTGGAAGWNISQNMGSALLPLTIALLTSM 175
 + ILL L G QG PA + +++ K RG ++ WN++ N+G L I + +
 Sbjct: 140 MFILLFLNGWFQGMGWACGRITMVHWSRKERGEIVSVWNVAHNVGGGL----IGPIFL 195

Query: 176 GLVVPANGNILLAFILPGVLVFLFALCCWKLGGDNPESEGLDSLRTMYGDAGESAVASEE 235
 GL + N + AF +P L A+ W + D P+S GL + D + S E
 Sbjct: 196 GLWM-FNDDWRTAFYVPAFFAVLVAVFTLVMRDTPQSCGLPPIEYKNDYPDDYDKSHE 254

Query: 236 EKHNLISYWLQIWKYVFCNPSSLLLVAANVVALYFVRFGIEDWMPYILSQVANMSEAHIFHA 295
 + ++ ++ +KYVF N L +A N +Y +R+G+ DW P+YL + + + +A
 Sbjct: 255 NE--MTAKEIFFKYVFNNKLLWSIATANAFVYLIRYGVLDWAPVYLKEAKHFTVDKSSWA 312

Query: 296 ISMLEWVAIPGSLVFAWLAVR-YPNKMAGVGAIGLFVLAIVFYERLTATGAPNYFLLL 354
 + EW IPG+L+ W++ + + + A G + + ++ V VY G P +
 Sbjct: 313 YFLYEWAGIPGTLGCGWISDKVFKGRRAPAGILFMVLVTLAVLVY-WFNPAGNPAVDMAA 371

Query: 355 VIAGILGSLIYGPQLIVNILTINFVPLNVAGTAIGFVGVTAYLIGNMGANWMLPILADGF 414
 ++A +G LIYGP +++ + + P AGTA G G+ YL G + AN ++ D F
 Sbjct: 372 LVA--IGFLIYGPVMLIGLYALELAPKKAAGTAAGLTGLFGYLGGAVAANAILGYTVDFH 429

Query: 415 GWFWSYIVVAALSAFSAV 432
 GW ++V+ A S +
 Sbjct: 430 GWDGGFMVLVASCVLVSL 447

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/439 (26%), Positives = 203/439 (45%), Gaps = 27/439 (6%)

Query: 23 KYPRYRVQVLISIFVGYMGYYFVRNNTSILSGILNMS----ATEIGIITCASYIAYGLSK 78
 ++ R + V F GY+ Y VRN ++S + + +I I+ ++YGL+K
 Sbjct: 17 QFLRRQKVFFFAFFGYVCAYLVRNNFKLMSNTIMVQNGWDKAQIAILLSCLTIVSYGLAK 76

-2351-

Query: 79 FISGLISDESNSKIFLPVGLFLTGLVNVLIGVIPSVITSIWLFITIMYLINGWLQGMGYPP 138
 F G + D + + + L + L+ +LIG S S+ + I+ ++ G +QG P
 Sbjct: 77 FYMGALGDRVSLRKLFSSISLGASALICILIGFFNS---SMVVLGILLVLGCVVQAGALAPA 133

5 Query: 139 GARTLVYWDYDNKERIKYATIWNLSHNFEGGAIAP I---LTGVGLALAGNDSLQARAAW 194
 + ++ NK R WN+S N G A+ P+ LT +GL + N ++ A+
 Sbjct: 134 SQAMIANYPNKRTRGGAIAGWNISQNMGSALLPLTIALLTSMGLVVPANGNI---LLAFL 190

10 Query: 195 FPGVVACLLAVLVYFLQEDTPESIGLPPIEEYHKEQYTNVVDSSDILEEPEVLGMGEIHK 254
 PGV+ L A+ + L D PES GL + + + + V S EE L ++I
 Sbjct: 191 IPGVLVFLFALCCWKLGGDNPESEGLDSLRTMYGDAGESAVASE---BEKHNL SYQLIW 247

15 Query: 255 KYILPNTKLMWASLYSIFVYILRYGIVSWTPKFLATSVDGGKGITATAGMGGFSLFEIG 314
 KY+ N L+ + ++ +Y +R+GI W P +L+ I S+ E
 Sbjct: 248 KYVFCNP SLLLVA AVNVALYFVRFGIEDWMPYILSQVANMSEAHIFHA-----ISMLEWV 302

20 Query: 315 GIIGMLTAGYLSAKVFNKSKPLTNVAFVLVAILLLAAYWFIPAG-PQYMALDFIILG-L 372
 I G L +L+ + + + V+A ++ G P Y L +++ G L
 Sbjct: 303 AIPGSLVFAWLAVRYPNKMAKVGALFLVLAIVFVYERLTATGAPNYFLL--LVIAGIL 360

Query: 373 GASTYGPVMMVGLYAMELVKAAAGASGLTGTFYSYVGATATLAIGI IIDHFGWGVAF 432
 G+ IYGP ++V + + VP AG A G G +Y+ G A + I+ D FGW ++
 Sbjct: 361 GS LIYGPQLIVNILTINFVPLNVAGTAIGFVGVTAYLIGNMGANWMLPILADGFGWFWSY 420

25 Query: 433 IIF-GISGFAAIVCTLLSR 450
 I+ +S F+A+ +L++
 Sbjct: 421 IVVAALSAFSAVGYLILAK 439

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 2083

A DNA sequence (GBSx2198) was identified in *S.agalactiae* <SEQ ID 6447> which encodes the amino acid sequence <SEQ ID 6448>. Analysis of this protein sequence reveals the following:

Possible site: 21
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3202(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6449> which encodes the amino acid sequence <SEQ ID 6450>. Analysis of this protein sequence reveals the following:

Possible site: 19
 45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4473(Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/100 (54%), Positives = 67/100 (67%)

55 Query: 1 MTYELCLEYGTYP LRPVDAWADEINTAPAFITEDKKLLELLEEVNTLFHFLFTIECSFH 60
 MTYELCLEYGTYP L VDA+ E P FI ED+ L LE +N LFH+LF+TIE FH
 Sbjct: 1 MTYELCLEYGTYP LSRVDAYWGEDQNPPTFIQEDRLCHKLETMNLPHDLFVTIESQFH 60

-2352-

Query: 61 YIGHDFPEKRAKITQIYHVIIIEHLSIHYPEYDIKIESLLM 100
 Y+G + PEKRA+I +Y + L Y +Y IKIE+ L+
 Sbjct: 61 YVGFNMPEKRAQIRILYQEVATILKSKYKDYPKIKIETPLL 100

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2084

A DNA sequence (GBSx2199) was identified in *S.agalactiae* <SEQ ID 6451> which encodes the amino acid sequence <SEQ ID 6452>. Analysis of this protein sequence reveals the following:

10 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2369(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAB81912 GB:U92974 unknown [Lactococcus lactis]
 Identities = 213/322 (66%), Positives = 260/322 (80%), Gaps = 5/322 (1%)

Query: 1 MSEKIRVLLYYKYVSIENAEYYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60
 M++ RVLLEY+YV IE+ E +A KHL CK +GLKGRIL+ADEGINGTVSG E T Y
 Sbjct: 1 MTQDYRVLLYYQYVPIEDGETFAQKHLADCKELGLKGRILVADEGINGTVSGTIEQTINAY 60

25 Query: 61 MDVHVSDFADLWFKIDEENQOAFKMFVRYKKEIVHLGLEDDNNFSDINPLETTGEYL 120
 M+ + +D RF+ FKIDE Q AF+KM VRY+ E+V+L LED D+NPLE TG YL
 Sbjct: 61 MELMKNDPRFSSTIFKIDEAEQNAPKKMHVRYRPELVNLSLED-----DVNPLELTGAYL 115

30 Query: 121 NPKQFKEALLDEDTVLDTRNDYEDLGHFGRGAIRPDIRNFRELQWVRDNKDKFMEKRV 180
 +PK+F+EA+LDE+TVV+D RNDYE+DLGHFRGAIRP+IR+FRELPQW+RDNK++FMEKRV
 Sbjct: 116 DPKEFREAML DENTVVIDARNDYEPDLGHFRGAIRPEIRSFRELQWIRDNKEQFMEKRV 175

35 Query: 181 VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGEGLWDGAMYVFD DRISVP 240
 + YCTGG+RCEKFSGW+VREGFKDVGQL GGIATYGKDPEVQG+LWDG MYVFD RI+VP
 Sbjct: 176 LTYCTGGIRCEKFSGWLVRGFKDVGQLLGGIATYGKDPEVQGD LWDGQMYVFD SRIAPV 235

40 Query: 241 INHVNPVISKDYFDGTPCERYVNCANPFCNKQIFASEENEAKYVRGCSPECRAHERNRY 300
 IN ++ +D+FDG+PCERY+NC NP CN+Q+ ASEENEAKY+ CS ECR H NRY
 Sbjct: 236 INQKEHVIVGRDWF DGSPCERYINCGNPECNRQMLASEENEAKYLGACSHCRVHPNNRY 295

Query: 301 VQENGLSRQEWAEERLEAIGESL 322
 ++ + LS QE ERL + + L
 Sbjct: 296 IKAHQLSNQEVQERLALLEKDL 317

45

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6453> which encodes the amino acid sequence <SEQ ID 6454>. Analysis of this protein sequence reveals the following:

50 Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.2443(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 321/324 (99%), Positives = 323/324 (99%)

Query: 1 MSEKIRVLLYYKYVSIENAEYYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60

-2353-

MSEKIRVLLYYKYVSIENA+EYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY
 Sbjct: 1 MSEKIRVLLYYKYVSIENAQEYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60
 Query: 61 MDWVHSDERFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDNNFSDINPLETTGEYL 120
 MDWVHSDERFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDNNFSDINPLETTGEYL
 Sbjct: 61 MDWVHSDERFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDNNFSDINPLETTGEYL 120
 Query: 121 NPKQFKEALLDEDTVLDTRNDYEDLGHFRGAIRPDIRNFRELPOQWVRDNKDKFMEKRV 180
 NPKQFKEALLDEDTVLDTRNDYEDLGHFRGAIRPDIRNFRELPOQWVRDNKDKFMEKRV
 Sbjct: 121 NPKQFKEALLDEDTVLDTRNDYEDLGHFRGAIRPDIRNFRELPOQWVRDNKDKFMEKRV 180
 Query: 181 VVYCTGGVRCCKFSGWMVREGFKDVGQLHGGIATYKGDPEVQOGELWDGAMYVFDDRISVP 240
 VVYCTGGVRCCKFSGWMVREGFKDVGQLHGGIATYKGDPEVQOGELWDGAMYVFDDRISVP
 Sbjct: 181 VVYCTGGVRCCKFSGWMVREGFKDVGQLHGGIATYKGDPEVQOGELWDGAMYVFDDRISVP 240
 Query: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENEAKYVRGCSPECRAHERNRY 300
 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENE KIVRGCSPECRAHERNRY
 Sbjct: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENETKYVRGCSPECRAHERNRY 300
 Query: 301 VQENGLSRQEWAEERLEAIGESLPQ 324
 VQENGLSRQEWAEERLEAIGESLP+
 Sbjct: 301 VQENGLSRQEWAEERLEAIGESLPE 324

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2085

A DNA sequence (GBSx2200) was identified in *S.agalactiae* <SEQ ID 6455> which encodes the amino acid sequence <SEQ ID 6456>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC83954 GB:L47648 putative [Bacillus subtilis]
 Identities = 54/192 (28%), Positives = 89/192 (46%), Gaps = 14/192 (7%)
 Query: 5 QTTIIAGAGAAGIGFGSAMQRLGLTNFLIIEKGHIGESFLRWPRTTQFITPSFTTNGFGFP 64
 + IIIIG G G+ ++++G+ + L+IEKG++ S +P F + S
 Sbjct: 5 KAIITGGGPGCLSAAILHLKQIGI-DALVIEKGNVNSIYNYPHTQTFSSSEKLE----- 58
 Query: 65 DLNAVIPDTPSPAFSFEKEHLSGVEYARYLQLVAAHYNLPIQNETSVLSIDK-RDSL FVIK 123
 I D A F E ++ Y + V N+ + V + K +++ FVI+
 Sbjct: 59 -----IGDV--AFITENRKPVRIQALSYYREVVKRKNIRVNAFEMVRKVTKTQNNTFVIE 111
 Query: 124 TSKGDFSADYLIMATGEFQNPNTIDIKGADLGMHYGQVDNFHIKSDNPFIIIGGNESACD 183
 TSK ++ Y I+ATG + +PN + + G DL + H D ++IGG S+ D
 Sbjct: 112 TSKETYTTTPYCIIATGYDHPNMGVPGEDLPKVFHYFKEGHPYFDKDVVIGGKNSSVD 171
 Query: 184 ALTHLVYLGQNV 195
 A LV G +V
 Sbjct: 172 AALELVKSGARV 183

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2354-

A related GBS gene <SEQ ID 8973> and protein <SEQ ID 8974> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1   Crend: 2
McG: Discrim Score:      5.05
5  GvH: Signal Score (-7.5): -3.14
    Possible site: 57
    >>> Seems to have an uncleavable N-term signal seq
ALOM program count: 0 value: 0.26 threshold: 0.0
    PERIPHERAL Likelihood = 0.26      6
10  modified ALOM score: -0.55

*** Reasoning Step: 3

----- Final Results -----
15      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

20      33.2/56.1% over 281aa

                                     Bacillus subtilis
EGAD|109228| hypothetical protein Insert characterized
GP|2635109|emb|CAB14605.1||Z99117 alternate gene name: yrdP Insert characterized
25  GP|1934657|gb|AAB80908.1||U93876 hypothetical protein YrdP Insert characterized
    PIR|E69725|E69725 potassium uptake trkA - Insert characterized

ORF01799(310 - 1128 of 1725)
EGAD|109228| S2656(2 - 283 of 345) hypothetical protein { acillus subtilis}
30  GP|2635109|emb|CA 14605.1||Z99117 alternate gene name: yrdP { acillus subtilis}
    GP|1934657|gb|AA 80908.1||U93876 hypothetical protein YrdP { acillus subtilis}
    PIR|E69725|E69725 potassium uptake trkA - acillus subtilis
    %Match = 6.1
    %Identity = 33.2 %Similarity = 56.0
35  Matches = 77 Mismatches = 88 Conservative Sub.s = 53

270      300      330      360      390      417      444      474
YYC*LVKYPILHIYFCQGEDMKHYQTIIIGAGAAGIGFGSAMQRLGLTNFLIIEKGH-IGESFL-RWPRTTQFITPSFTT
40      |||:|||| ||| | ::: |:::| | :||| : | : :
      MYDTIVIGAGQAGISIGYYLKQ-SDQKFIILDKSHEVGESWKDRYDSLVLFTSRMYSS
      10      20      30      40      50

480      510      540      570      600      630      660      690
-----NGFGFPDLNAVIPDTSPAFSFEKEHLSGVEYARYQLVAAHYNLPIONETSVLSIDKRDSLFIKTSKGFDS
45      ||| | : : ||: : : ||| | | : | : | : | : :
LPGMHLEGEKKGFPFSKNEIV-----AYLKKYVKKFETPIQLRTEVISVLKIKNYFLIKTNREEYQ
      70      80      90      100      110

720      750      822      852      882      912
ADYLIMATGEFQNPNTIDIKGADLG-----MHYQVDNF-HIKSDNPFIIIGGNESACDALTHLVYLGQVELYTDTFGR
50      |::||| | : || | || : | | | : : | : : ||| |
TKNLVIATGPFHFTFNIPSIS-KDLSDNINQLHSSQYKNSQLAYGNVLVVGCGNSGA-----
      130      140      150      160      170

942      969      996      1026
KESNPDPSISLS-PLTKERLKHQ-DHKKEYYSISEGKKAI--EIKQIG-----
60      :: |::| : : : | : : | : : | : :
-----QIAVELSKERVTYLACSNKLVYFPLMIGKRSIFWWFDKLGVLHASHTSIVGKFIQKGDVPVFGHELKHAIK
      180      190      200      210      220      230      240

1068      1098      1128      1158      1188      1218      1248
-----KQYQVTFDDGSTAESFHKPILSTGFLNTCHLIDGIALFEYDKNQLPIVTEDESTIVNNCFLIGPSL
65      || : : | | | | : : | : ||| | | : : : : : :
QKEIILKKRVIAAKQNEIIFKDSSTLE-VNNIIWATGFRNPLCWINIKGVLDQEGRIIHHRGVSPVEGLYFGLPWQHKR
      260      270      280      290      300      310      320

```

-2355-

SEQ ID 8974 (GBS284) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 10; MW 42.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 9; MW 67.6kDa).

GBS284-GST was purified as shown in Figure 225, lane 7.

5 Example 2086

A DNA sequence (GBSx2201) was identified in *S.agalactiae* <SEQ ID 6457> which encodes the amino acid sequence <SEQ ID 6458>. This protein is predicted to be NrgA-like protein. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
  INTEGRAL    Likelihood = -11.73    Transmembrane    7 - 23 ( 1 - 31)
  INTEGRAL    Likelihood = -6.42     Transmembrane    86 - 102 ( 82 - 108)
  INTEGRAL    Likelihood = -6.42     Transmembrane    324 - 340 ( 318 - 342)
  INTEGRAL    Likelihood = -5.26     Transmembrane    210 - 226 ( 207 - 229)
15  INTEGRAL    Likelihood = -5.10     Transmembrane    113 - 129 ( 112 - 133)
  INTEGRAL    Likelihood = -1.49     Transmembrane    246 - 262 ( 246 - 263)
  INTEGRAL    Likelihood = -1.17     Transmembrane    183 - 199 ( 183 - 199)
  INTEGRAL    Likelihood = -0.43     Transmembrane    41 - 57 ( 41 - 57)
20  INTEGRAL    Likelihood = -0.00     Transmembrane    265 - 281 ( 265 - 282)

----- Final Results -----
      bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9997> which encodes amino acid sequence <SEQ ID 9998> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15668 GB:Z99122 ammonium transporter [Bacillus subtilis]
30  Identities = 105/378 (27%), Positives = 181/378 (47%), Gaps = 41/378 (10%)

Query: 3  VKKGLFVFLLLCILSMWLMIFGVAFYFYFGLH-QSLTSRIIYQFVLTVLLTTTAWFMGAY 61
      ++ G VF+ C L +WLM G+A +Y G + +++ S ++ F ++ + + W + Y
Sbjct: 1  MQMGDTVMFFCALLVWLMTPGLALFYGGMVKSKNVLSTAMHSFS-SIAIVSIVWVLFY 59

35  Query: 62 FLAFEGHFKTVQFQEQEADGKQI-----VNCLFQLCFALYAVVMLIGSIIDR 107
      LAF + + A K + + +FQ+ FA+ ++ G+ +R
Sbjct: 60 TLAFAPGNSIIIGLEWAGLKGVGFDPGDYSDTIPHSIFMMFQMTFAVLTTAIIISGAFAR 119

40  Query: 108 VQTKRLLAVVSWLFLVYTPPLAYLIWNSEGVFVAKMGVLDVSGGMIVHLSAGLSSYILAHV 167
      ++ LL V W LVYTP+A+ +W G ++G LDF+GG +VH+S+G++ +LA V
Sbjct: 120 MRFGAFLLFSVLWASLVYTPVAHWVWGG-GWIGQLGALDFAGGNVVHISGVAGLVLAIV 178

45  Query: 168 IGK-----SEHQHNKVKNDLFLGLMILITFGWFGFNMGVGEWNSQAIMILLNTIFAIG 222
      +GK + HN + FLG LI FGWFGFN+G + A+ +NT A
Sbjct: 179 LGKRKDGTAASSPHNLIYT---FLGGALIWFVGFGFVGSALTLDGVAMYAFINTNTAAAA 235

Query: 223 GGLAWTLAAKWNGEEEKTGSLLNGIIVGLVTSTAGVGYLLTWQLLAVTFFASLFTYFVTD 282
      G W L ++ ++G I GLV T G++ + + + ++
50  Sbjct: 236 GIAGWILVEWIINKKPTMLGAVSGALAGLVAITPAAGFVTPPFASIIIGIIGAVCFWGVF 295

Query: 283 YVAKAFAIDDVVSSFGMNGIGGLLGLSLGVGLFKLSHMP-----VQLLAL 326
      + K F DD + +FG++GIGG G + GLF + + Q++A+
Sbjct: 296 SLKKKFGYDDALDAFGLHGIGGTWGGIATGLFATTSVNSAGADGLFYGDASLIWKQIVAI 355

55  Query: 327 ATTILLSIIMTYIISKAI 344
      A T + I+T++I K +
Sbjct: 356 AATYVFVFIVTFVVIKIV 373

```

-2356-

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8975> and protein <SEQ ID 8976> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: -1   Crend: 4
      McG: Discrim Score:      17.19
      GvH: Signal Score (-7.5): -4.07
      Possible site: 24
      >>> Seems to have an uncleavable N-term signal seq
10     ALOM program      count: 9 value: -11.73 threshold: 0.0
      INTEGRAL      Likelihood = -11.73   Transmembrane      7 - 23 ( 1 - 31)
      INTEGRAL      Likelihood = -6.42   Transmembrane      86 - 102 ( 82 - 108)
      INTEGRAL      Likelihood = -6.42   Transmembrane     324 - 340 ( 318 - 342)
      INTEGRAL      Likelihood = -5.26   Transmembrane     210 - 226 ( 207 - 229)
15     INTEGRAL      Likelihood = -5.10   Transmembrane     113 - 129 ( 112 - 133)
      INTEGRAL      Likelihood = -1.49   Transmembrane     246 - 262 ( 246 - 263)
      INTEGRAL      Likelihood = -1.17   Transmembrane     183 - 199 ( 183 - 199)
      INTEGRAL      Likelihood = -0.43   Transmembrane      41 - 57 ( 41 - 57)
      INTEGRAL      Likelihood = -0.00   Transmembrane     265 - 281 ( 265 - 282)
20     PERIPHERAL      Likelihood = 0.26      152
      modified ALOM score: 2.85

      *** Reasoning Step: 3

25     ----- Final Results -----
      bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30     The protein has homology with the following sequences in the databases:

      ORF01800(307 - 1332 of 1641)
      EGAD|19589|BS3646(1 - 373 of 404) probable ammonium transporter {Bacillus subtilis}
      OMNI|NT01BS4254 ammonium transporter SP|Q07429|NRGA_BACSU PROBABLE AMMONIUM TRANSPORTER
      (MEMBRANE PROTEIN NRGA). GP|143264|gb|AAA17399.1||L03216 membrane-associated protein
35     {Bacillus subtilis} GP|1684645|emb|CAB05374.1||Z82987 unknown {Bacillus subtilis}
      GP|2636176|emb|CAB15668.1||Z99122 ammonium transporter {Bacillus subtilis}
      PIR|A36865|A36865 ammonium transporter nrgA - Bacillus subtilis
      %Match = 13.5
      %Identity = 30.0 %Similarity = 54.8
40     Matches = 104 Mismatches = 149 Conservative Sub.s = 86

      144      174      204      234      264      294      324      354
      PFSMIRKVFSPNRCMAEPKPIPAAPAPIIMV**CFMSSP*QK*MCKIKYLT*S*Q*YSLTNKRVFVKKGLFVFLLLCILSM
      :: | ||:::| | :
45     MQMGDTVFMMFFCALLV
      10

      384      411      441      471      501      531
      WLMIEGVAFYYFGSLH-QSLTSRIIYQFVLTVLLTTTAWFMGAYFLAFEGHFKTVFQFQEADGKQI-----
50     ||| |:|::| | : ::| ::| ::| :| :| ||| : ::| | :
      WLMTFGLALFYGMVKSKNVLSTAMHSF-SSIAIVSVVWLFEGYTLAFAPGNSIIGGLEWAGLKGVGFDPGDYSDTIPHS
      30      40      50      60      70      80      90

      579      609      639      669      699      729      759      789
      VNCLFQLCFALYAVVMLIGSIIDRVQTKRLLLA VSWLFLVYTPPLAYLIWNSEGVFAKMGVLD FSGGMIVHLSAGLSSYI
55     : ::| :| : ::| :| : ::| :| :| |||::|:| :| :| |||::|:| :| :
      LFMMFQMTFAVLTTAIIISGAFARMRFGAFLLFPSVLWASLVYTPVAHWVWGG-GWIGQLGALDFAGGNVHHISSGVAGLV
      110      120      130      140      150      160      170

      819      849      873      903      933      963      993      1023
      LAHVIGKSEHQHNKVKNDSLF--LGMILITFGWFGFMGPGVEWNSQAIMILLNTIFAIIGGLAWTLA AKWNGEEETG
60     || |:| : ::| :| || |||||::| : |: ::| | | | :| :|
      LAIVLGKRKDGTAASSPHNLIYTFLLGALIWFGWFGFNVGSALTLDGVAMYAFINTINTAAAAGIAGWIL-VEWIINKKPTM
      190      200      210      220      230      240      250

```

-2357-

```

1050      1080      1110      1140      1170      1200      1230      1260
-SLLNGIIVGLVLTSTAGVGYLLTWQLLAVTFFASLFTYFVTDYKAKAFAIDDVSSFGMNGIGLLGSLGVGLFKLSHMP
  ::| | || | | :: : : : : : : | | : : ||::||| | : ||| : :
5 LGAVSGAIAGLVAITPAAGFVTPFASIIIGIIGGAVCFWGVFSLKKKFGYDDALDAFGLHGIGGTWGGIATGLFATTSVN
      270      280      290      300      310      320      330

1272      1302      1332      1362      1392      1422      1452
V-----QLLALATTILLSIIMTYIISKAFRK**IRLRCTSQPYLLF*QGE*LNRIINHFIHY*TLSXX*
  ::||:| | :: |::| | :
10 SAGADGLFYGDASLIWKQIVAIATYVFVFIIVTFVIKIVSLFLPLRATEEEEEESLGLDLTMHGEKAYQDSM
      350      360      370      380      390      400

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 15 vaccines or diagnostics.

Example 2087

A DNA sequence (GBSx2202) was identified in *S.galactiae* <SEQ ID 6459> which encodes the amino acid sequence <SEQ ID 6460>. This protein is predicted to be dUTPase (dut). Analysis of this protein sequence reveals the following:

```

20 Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2731(Affirmative) < succ>
25    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9471> which encodes amino acid sequence <SEQ ID 9472> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA72644 GB:Y11901 dUTPase [Lactococcus lactis]
  Identities = 67/144 (46%), Positives = 90/144 (61%), Gaps = 8/144 (5%)

Query: 40  RGFELVSQFNSKELLPKRETAHAAGYDLKVAKKTVIEPGEITLVPTGIKAYMQPGEVLVL 99
35  RGF+   +      +P+R T H+AGYD+  ++   I+P EI +V TG+   +   EVL L
Sbjct: 3   RGFK---KLDGNATIPERATKHSAGYDISASEVTIQQPDEIKMVSTGLAVQLGDDEVKLK 59

Query: 100 YDRSSNP RKGI V LINSVGVIDG DYNNQVNEGHIFAQM QNITDQAVILEEGERIVQAVF 159
40  YDRSSNP K+GI LINSVG+ID DYY   +      NI+ + V + +G+RI+Q VF
Sbjct: 60  YDRSSNPVKRGIALINSVGIIIDSYYPQEFK-----GLFMNISKEPVTISKGORIMQGVF 114

Query: 160 APFL LADDDQATGMRTGGFGSTGK 183
      +L DDD A G RTGGFGSTG+
45 Sbjct: 115 VKYL TIDDDNANGKRTGGFGSTGE 138

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6461> which encodes the amino acid sequence <SEQ ID 6462>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2519(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-2358-

Identities = 115/148 (77%), Positives = 125/148 (83%)

Query: 36 MSKVRGFELVSQFSNKELLPKRETAHAAGYDLKVAKKTVEIEPGEITLVPTGIKAYMQPGE 95
 M+K+RGFELVS F+N +LLPKRET HAAGYDL VA+ I PGEI LVPTG+KAYMQ GE
 Sbjet: 1 MTKIRGFELVSSFTNPDLPLPKRETTAAGYDLSVAEAVTIAPGEIKLVPTGVKAYMQDGE 60

Query: 96 VLYLYDRSSNPRKKGIVLINSVGVIDGDYNNQVNEGHIQAQMNITDQAVILEEGERIV 155
 VLYLYDRSSNPRKKGII+LINSVGVID DYY N+ NEGHIQAQMNITD V L GERIV
 Sbjet: 61 VLYLYDRSSNPRKKGIIILINSVGVIDADYNGNEANEGHIQAQMNITDHPVTLAVGERIV 120

Query: 156 QAVFAPFLADDDQATGMRTGGFGSTGK 183
 Q VF PFL+AD DQA G RTGGFGSTG+
 Sbjet: 121 QGVFMPFLIADGDQARGERTGGFGSTGQ 148

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2088

A DNA sequence (GBSx2203) was identified in *Sagalactiae* <SEQ ID 6463> which encodes the amino acid sequence <SEQ ID 6464>. This protein is predicted to be RadA homolog (radA). Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2628(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11863 GB:Z99104 DNA repair protein homolog [Bacillus subtilis]
 Identities = 285/453 (62%), Positives = 358/453 (78%), Gaps = 4/453 (0%)

Query: 1 MAKKKSVFTCQECGYQSPKYLGRCPNCSAWSSFFVEEVEVQEVKNARVSLNGEKSRPTKLK 60
 MAK KS F CQ CGY+SPK++G+CP C AW++ VEE+ + N R + + K
 Sbjet: 1 MAKTKSKFICQSCGYESPKWMGKCPGCGAWNTMVEEMIKKAPANRRAAFSSHVQTVQKPS 60

Query: 61 DVSSINYS---RTKTDMEFNRLVGGGVVPGSLVLIGGDPGIGKSTLLLQVSTQLA-NKG 116
 ++SI S R KT + EFNRLVGGGVV GSLVLIGGDPGIGKSTLLLQVS QL+ +
 Sbjet: 61 PITSIETSEEPVRVKTQLGEFNRLVGGGVVKGSLVLIGGDPGIGKSTLLLQVSAQLSGSSN 120

Query: 117 TVLYVSGEESAEQIKLRSERLGDIDNEFYLYAETNMQSIRSEIEKIKPDLFIIDSQITIM 176
 +VLY+SGEES +Q KLR++RLG + ++ +ET+M+ I S I+++ P F+++DSIQT+
 Sbjet: 121 SVLYISGEESVKQTKLRADRLGINNPSLHVLSETDMEYISSAIQEMNPSFVVVDSIQTVY 180

Query: 177 SPEVSSVQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKEGTLAGPRMLEHMDTVLYF 236
 +++S GSVSQVRE TAELM++AKT I FIVGHVTKEG++AGPR+LEHMDTVLYF
 Sbjet: 181 QSDITSAPGSVSQVRECTAELMKIAKTGIPFIVGHVTKEGSIAGPRLLLEHMDTVLYF 240

Query: 237 EGERHHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSAIVVT 296
 EGERHHTFRILRAVKNRFGSTNE+GIFEM+ GL EVLNPS++FLEER G+ GS+I +
 Sbjet: 241 EGERHHTFRILRAVKNRFGSTNEMGIFEMREEGLTEVLNPSFIFLEERSAGSAGSSITAS 300

Query: 297 MEGTRPILAEVQALVPTPTVFGNAKRITTTGLDFNRVSLIMAVLEKRCGLLLQNDAYLKSA 356
 MEGTRPIL E+QAL++PT FGN +R TG+D NRVS+MAVLEKR GLLLLQNDAYLK A
 Sbjet: 301 MEGTRPILVEIQALISPTSFNGNRRMATGIDHNRVSLIMAVLEKRVGLLLQNDAYLKVA 360

Query: 357 GGVKLDEPAIDLAVAVAIASSYKEKPTNPQESFIGEIGLTGEIRRVTRIEQRINEASKLG 416
 GGVKLDEPAIDLA+ ++IASS+++ P NP + FIGE+GLTGE+RRV+RIEQR+ EA+KLG
 Sbjet: 361 GGVKLDEPAIDLAIVISIASFRDTPPNPADCFIGEGLTGEVRRVSRIEQRVKEAAKLG 420

Query: 417 FTKIYAPKNSLAGIEIPKGIDVIGVTTVSQVLK 449

-2359-

F ++ P +L G PKGI+VIGV V++ L+
 Sbjct: 421 FKRMIIPAANLDGWTCPKGIEVIGVANVAEALR 453

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6465> which encodes the amino acid sequence <SEQ ID 6466>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2191(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 416/453 (91%), Positives = 441/453 (96%)

Query: 1 MAKKKSVFTCQECGYQSPKYLGRCPNCSAWSSFVEEVEVQEVKNARVSLNGEKS RPTK LK 60
 MAKKK+ F CQECGYQSPKYLGRCPNCSAWSSFVEEVEV+EVKNARVSL GEKSRP KLK
 Sbjct: 1 MAKKKATFICQECGYQSPKYLGRCPNCSAWSSFVEEVEVKEVKNARVSLAGEKSRPVKLK 60

Query: 61 DVSSINYSRTKTDMDDEFNRVLGGGVVPGSLVLIGGDPGIGKSTLLQVSTQLANKGTVLY 120
 DV +I+Y RT+TDM EFNRLVGGGVVPGSL+LIGGDPGIGKSTLLQVSTQLANKGTVLY
 Sbjct: 61 DVDNISYHRTQTDMSEFNRLVGGGVVPGSLVLIGGDPGIGKSTLLQVSTQLANKGTVLY 120

Query: 121 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQSIRSEIEKIKPDFLIIDSIQTIMSPEV 180
 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQ+IR+EIE IKPDFLIIDSIQTIMSP++
 Sbjct: 121 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQAIRTEIENIKPDFLIIDSIQTIMSPDI 180

Query: 181 SSVQGSVSQVREVTAEMLQAKTNNIATFIVGHVTKEGTLAGPRMLEHMDTVLYFEGE 240
 + VQGSVSQVREVTAEMLQAKTNNIATFIVGHVTKEGTLAGPRMLEHMDTVLYFEGE
 Sbjct: 181 TGVQGSVSQVREVTAEMLQAKTNNIATFIVGHVTKEGTLAGPRMLEHMDTVLYFEGE 240

Query: 241 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSAIVVTMEGT 300
 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSA+VVTMEGT+
 Sbjct: 241 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSAVVVTMEGS 300

Query: 301 RPILAEVQALVPTPTVFGNAKRTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK 360
 RPILAEVQ+LVTPTPTVFGNA+RTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK
 Sbjct: 301 RPILAEVQSLVPTPTVFGNARRTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK 360

Query: 361 LDEPAIDLAVAVAIASSYKEKPTNPQESFIGEIGLTGEIRRVTRIEQRINEASKLGFTKI 420
 LDEPAIDLAVAVAIASSYKEKPT+PQE+P+GEIGLTGEIRRVTRIEQRINEA+KLGFTK+
 Sbjct: 361 LDEPAIDLAVAVAIASSYKEKPTSPQEAFLGEIGLTGEIRRVTRIEQRINEAAKLGFTKV 420

Query: 421 YAPKNSLAGIEIPKGIDVIGVTTVVSQVLKAVFS 453
 YAPKN+L GI+IP+GI+V+GVTTV QVL AVFS
 Sbjct: 421 YAPKNALQGDIPQGIIEVVGVTTVGQVLNAVFS 453

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2089

A DNA sequence (GBSx2204) was identified in *S.agalactiae* <SEQ ID 6467> which encodes the amino acid sequence <SEQ ID 6468>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2360-

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAA97750 GB:Z73419 hypothetical protein Rv1284 [Mycobacterium
    tuberculosis]
    Identities = 69/162 (42%), Positives = 100/162 (61%), Gaps = 2/162 (1%)

    Query: 3  TYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRNAG 62
              T D++L N YA      LP+ P  +AIV CMD+RL V + LG+  G+AH++RNAG
    10  Sbjct: 2  TVTDDYLANNVDYASGF-KGPLPMPSPSKHIAIVACMDARLDVYRMLGIKEGEAHVIRNAG 60

    Query: 63  GRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTF'TNEAFAAQLQRD LGVDMHGHDFLP 122
              VTDDV+RSL ISQ+ LGTREI++LHHTDCG TFT++ F  +Q + G+
    15  Sbjct: 61  CVVTDDVIRSLAISQRLGTREIILLHHTDCGMLTFTDDDFKRAIQDETGIRPTWSP-ES 119

    Query: 123 FNDIEESVREDVAKLHASPLIPDDVVISGAIYDVDTG RMVEV 164
              + D E VR+ + ++ +P +      + G ++DV TG++ EV
    15  Sbjct: 120 YPDAVEDVRQSLRRIEVNPFVTKHTSLRGFVFDVATGKLINEV 161
  
```

There is also homology to SEQ ID 6470:

```

20  Identities = 126/164 (76%), Positives = 146/164 (88%)

    Query: 1  MTTYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60
              + +YF++F+ NQAY LHGTAHL+KPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN
    25  Sbjct: 1  LMSYFEHFMAANQAYVALHGTAHLPLKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60

    Query: 61  AGGRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTF'TNEAFAAQLQRD LGVDMHGHDF 120
              AGGRVT+D++RSLVISQQQ+GTREIVVLHHTDCGAQTF'TNE FA  +  LGVD+ G DF
    30  Sbjct: 61  AGGRVTEDMIRSLVISQQQMG'TREIVVLHHTDCGAQTF'TNEGFAKHIHEHLGVDVSGQDF 120

    Query: 121 LPFNDIEESVREDVAKLHASPLIPDDVVISGAIYDVDTG RMVEV 164
              LPF D+E+SVRED+AK+ AS LI DDVVI+GA+YDVTG+M +V
    30  Sbjct: 121 LPFQDVEDSVREDMAKIRASSLISDDVINGAVYDVTGKMTQV 164
  
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2090

A DNA sequence (GBSx2205) was identified in *Sagalactiae* <SEQ ID 6471> which encodes the amino acid sequence <SEQ ID 6472>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 19
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.0536(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

A related GBS nucleic acid sequence <SEQ ID 9473> which encodes amino acid sequence <SEQ ID 9474> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

50  >GP:AAC73407 GB:AE000137 putative oxidoreductase [Escherichia coli K12]
    Identities = 199/438 (45%), Positives = 286/438 (64%)

    Query: 1  MKKYDVIVLFGFGKAGKTLAAKLATQGKSVAMVEEDDKMYGGTCINIGCIPTKTLVLSASK 60
              M KY  +++GFGKAGKTLA LA G VA++E+ + MYGGTCINIGCIPTKTL+ A +
    55  Sbjct: 10  MNKYQAVIIGFGKAGKTLAVTLAKAGWRVALIEQSNAMYGGTCINIGCIPTKTLVHDAQQ 69

    Query: 61  NHDFQEAMTTRNEVTSRLRAKNFAMLDNKTVDVYNAKARFISNKVVELTGADKQELTA 120
              + DF A+ +NEV + LR KNF L +  +DV + +A FI+N  + +  E+
  
```

-2361-

Sbjct: 70 HTDFVRAIQRKNEVVNFLRNKNFHNLAADMPNIDVIDGQABFINNHSRLVRHREPNLEIHG 129
 Query: 121 DVIIINTGAKSVQLPIPLGLADSQHVYDSTAIQELAHLPKRLGIIGGGNIGLEFATLYSEL 180
 + I INTGA++V PIPG+ + VYDST + L LP LGI+GGG IG+EFA++++
 Sbjct: 130 EKIFINTGAQTVVPPPIPGITTTTPGVYDSTGLLNLKELPGHLGILGGGYIGVRFASMPANF 189
 Query: 181 GSKVTVIDSQSRIFAREEEELSEMAQDYLEEMGISFKLSADIKSVQNEDEDVVISFEDEK 240
 GSKVT++++ S RE+ ++++ L + G+ L+A ++ + + + V + E +
 Sbjct: 190 GSKVTILEAASLFLPREDRDIADNIATILRDQGVDTILNAHVERISHHENQVQVHSEHAQ 249
 Query: 241 LSFDAVLATGRKPNTEGLALENTDIKLTERGAIADVDEYCQTSVENIFAVGDVNGGPQFT 300
 L+ DA+L A+GR+P T L EN I + ERGAI VD+ T+ +NI+A+GDV GG QFT
 Sbjct: 250 LAVDALLIASGRQPATASLHPENAGIAVNERGAIVVDKRLHTTADNIWAMGDVTGGLQFT 309
 Query: 301 YISLDDSRIVLNYLNCDDKDYSLKNRCVAPTSTFTNPPLATVGLDEKTAKEKGYQVKNSL 360
 YISLDD RIV + L + S +R VP S F PPL+ VG+ E+ A+E G ++ +L
 Sbjct: 310 YISLDDYRIVRDELLGEGKRSTDDRKNVPYSVFMTPPLSRVGMTEEQARESAGADIQVVTL 369
 Query: 361 LVSAMPRAHVNNDLRGIFKVVVDTETNLILGARLFGAESHELINIITMAMDNKIPYTYFQ 420
 V+A+PRA V ND RG+ K +VD +T +LGA L +SHE+INI+ M MD +PY+ +
 Sbjct: 370 PVAaipRARVMNDTRGVLKAIVDNKTQRLMGASLLCVDSEMINIVKMVMDAGLPYSILR 429
 Query: 421 KQIFTHPTMVENFNDLFN 438
 QIFTHP+M E+ NDLF+
 Sbjct: 430 DQIFTHPSMESLNDLFS 447

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2091

A DNA sequence (GBSx2206) was identified in *S. agalactiae* <SEQ ID 6473> which encodes the amino acid sequence <SEQ ID 6474>. This protein is predicted to be glutamyl-tRNA synthetase (gltX). Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2245(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9475> which encodes amino acid sequence <SEQ ID 9476> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10953> which encodes amino acid sequence <SEQ ID 10954> was also identified.

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC31971 GB:U49789 glutamyl-tRNA synthetase [Bacillus subtilis]
 Identities = 273/491 (55%), Positives = 353/491 (71%), Gaps = 19/491 (3%)
 Query: 20 LANKIRVRYAPSPTGLLHIGNARTALFNLYLARHHGGDFVIRIEDTDRKRHVEDGERSQL 79
 + N++RVRYAPSPTG LHIGNARTALFNLY+AR+ GG F+IR+EDTD+KR++E GE+SQL
 Sbjct: 1 MGNEVRVRYAPSPTGHLHIGNARTALFNLYFARNQGGKFIIRVEDTDKKRNIEGGEQSQL 60
 Query: 80 ENLRWLGMWDWDESPET---HENYRQSERLELYQRYIDQLLAEGKAYKSYVTEELAEERE 136
 L+WLG+DWDES + + YRQSER ++Y+ Y ++LL +G AYK Y TEEEL ERE
 Sbjct: 61 NYLKWLGIDWDESVDVGGEYGPYRQSERNDIYKVYYEELLEKGLAYKCYCTEELEKERE 120
 Query: 137 RQELAGETPRYINEFIGMSETEKEAYIAEREAAGIIPTVRLAVNESGIYKWTDMVKGDIE 196
 Q GE PRY + +++ E+E +IAE G P++R V E + + D+VKG+I

-2362-

5 Sbjct: 121 EQIARGEMPRYSGKHRDLTQEEQEKFIAB---GRKPSIRFRVPEGKVI AFNDIVKGEIS 176
 Query: 197 FE GSNIGGDWVIQKKGYPYTNFAVVIDDHMQISHVIRGDDHIANTPKQLMVYEALGWE 256
 FE IG D+VI KKD G PTYNFAV IDD+ M+++HV+RG+DHI+NTPKQ+M+Y+A GW+
 10 Sbjct: 177 FESDGIG-DFVIVKKDGTPTYNFAVAIDYLMKMT HVLRGEDHISNTPKQIMIQAFGWD 235
 Query: 257 APQFGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYMSEAVFNFIALLGWNPGGEEI 316
 PQFGHMTLI+N E+ KKL SKRD + +QFIE Y++ GY+ EA+FNFI LLGW+P GEEE+
 15 Sbjct: 236 IPQFGHMTLIVN-ESRKKLSKRDESIIQFIEQYKELGYLPEALFNFIGLLGWSVPVGEEL 294
 Query: 317 FSREQLINLFDENRLSKSPA AFDQKKMDWMSNDYLNADFSVFALCKPFL EEA GRL--- 373
 F++EQ I +FD NRLSKSPA FD K+ W++N Y+K D + V L P L++AG++
 20 Sbjct: 295 FTKEQFIEIFDVNRLSKSPALFDMHKLKWNQYVKKLDLDQVVELTLP H LKAGKVGTE 354
 Query: 374 -----TDKAEKLVELYQPOLKSADEIVPLTDLFFADFP ELTEAEKEVMAAETVPTVLSAF 428
 + KL+ LY QL EIV LTDLFF D E + K V+ E VP VLS F
 25 Sbjct: 355 LSAAEQEWVRKLISLYHEQLSYGAEIVELTDLFFTEIEYNQEAKAVLEEEQVPEVLSTF 414
 Query: 429 KEKLVSLSD EEFTRDTIFPQIKAVQKETGIGKKNLFMPIRIAVSGEMHGPELPDTIYLLG 488
 KL L EEFT D I IKAVQKETG KGK LFMPIR+AV+G+ HGPELP +I L+G
 30 Sbjct: 415 AAKLEEL--EEFTPDNIKASIKAVQKETG H K G K L F M P I R V A V I G Q T H G P E L P Q S I E L I G 472
 Query: 489 KEKSVQHIDNM 499
 KE ++Q + N+
 35 Sbjct: 473 KETAIQRLKNI 483

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6475> which encodes the amino acid sequence <SEQ ID 6476>. Analysis of this protein sequence reveals the following:

30 Possible site: 24
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1966 (Affirmative) < succ>
 35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 434/481 (90%), Positives = 459/481 (95%)

40 Query: 20 LANKIRVRYAPSPTGLLHIGNARTALFNLYARHHGGDFVIRIEDTDRKRHVEDGERSQL 79
 ++ IRVRYAPSPTGLLHIGNARTALFNLYAR HGG F+IRIEDTDRKRHVEDGERSQL
 Sbjct: 1 MSKPIRVRYAPSPTGLLHIGNARTALFNLYARRHGGTFIIRIEDTDRKRHVEDGERSQL 60
 45 Query: 80 ENLRWLGMWD E SPETHENYRQSERLELYQRYIDQLLAEGKAYKSYVTEEE LA AERERQE 139
 ENL+WLGMWD E SPETHENYRQSERL LYQ+YIDQLLAEGKAYKSYVTEEE LA AERERQE
 Sbjct: 61 ENLKLWGMWD E SPETHENYRQSERLALYQQYIDQLLAEGKAYKSYVTEEE LA AERERQE 120
 50 Query: 140 LAGETPRYINEFIGMSETEKEAYIAEREAAGIIPTVRLAVNESGIYKWTDMVKGDIEFEG 199
 AGETPRYINEFIGMS EK YIAEREAAGI+PTVRLAVNESGIYKWTDMVKGDIEFEG
 Sbjct: 121 AAGETPRYINEFIGMSADEKAKYIAEREAAGIVPTVRLAVNESGIYKWTDMVKGDIEFEG 180
 55 Query: 200 SNIGGDWVIQKKGYPYTNFAVVIDDHMQISHVIRGDDHIANTPKQLMVYEALGWEAPQ 259
 NIGGDWVIQKKGYPYTNFAVV+DDHMQISHVIRGDDHIANTPKQLMVYEALGWEAP+
 Sbjct: 181 GNIGGDWVIQKKGYPYTNFAVVDDHMQISHVIRGDDHIANTPKQLMVYEALGWEAPE 240
 60 Query: 260 FGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYMSEAVFNFIALLGWNPGGEEI FSR 319
 FGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYM EAVFNFIALLGWNPGGEEI FSR
 Sbjct: 241 FGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYMPEAVFNFIALLGWNPGGEEI FSR 300
 65 Query: 320 EQLINLFDENRLSKSPA AFDQKKMDWMSNDYLNADFSVFALCKPFL EEA GRLTDKAEK 379
 EQLI LFDENRLSKSPA AFDQKKMDWMSN+YLK+ADFE+V+ALCKPFL EEA GRLT+KAEK
 Sbjct: 301 EQLIALFDENRLSKSPA AFDQKKMDWMSNEYLKHADFETVYALCKPFL EEA GRLTEKAEK 360
 Query: 380 LVELYQPOLKSADEIVPLTDLFFADFP ELTEAEKEVMAAETVPTVLSAFKEKLVSLSD E 439
 LVELY+PQLKSADEI+PLTDLFF+DFPELTEAEKEVMA ETV TVL AFK KL ++SDE+

-2363-

Sbjct: 361 LVELYKPOLKSADEIIPLTDLFFSDFPELTEAEKEVMAGETVSTVLQAFKAKLEAMSDDED 420

Query: 440 FTRDTIFPQIKAVQKETGIKGNLFMPIRIAVSGEMHGPELPTIYLLGKEKSVQHIDNML 500
F + IFPQIKAVQKETGIKGNLFMPIRIAVSGEMHGPELP+TIYLLG++KS++HI NML

5 Sbjct: 421 FKPENIFFPQIKAVQKETGIKGNLFMPIRIAVSGEMHGPELPNTIYLLGRDKSIEHIKNML 481

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2092

10 A DNA sequence (GBSx2207) was identified in *S.agalactiae* <SEQ ID 6477> which encodes the amino acid sequence <SEQ ID 6478>. This protein is predicted to be d-ribose-binding protein precursor , fragment (rbsB). Analysis of this protein sequence reveals the following:

Possible site: 24

>>> May be a lipoprotein

15

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15613 GB:Z99122 ribose ABC transporter (ribose-binding protein) [Bacillus subtilis]

Identities = 143/301 (47%), Positives = 205/301 (67%), Gaps = 1/301 (0%)

25

Query: 14 MSIVILILGACCKTGLGNSSGNSTKNVTKKSAKDLKLGVSISTNNPYFVAMKDGIDKYAS 73
+S++L L T K + K+ +G+S+ST NNP+PV++K GI+K A

Sbjct: 5 VSVILTSLFLLTACSLLEPPQWAKPSNSGNKKEFTIGLSVSTLNNPFFVSLKKGIEKEAK 64

30

Query: 74 NKKISIKVADAQDDAARQADDVQNFISQNVDAILINPVDSKAIVTAIKSANNANIPVILM 133
+ + + + DAQ+D+++Q DV++ I Q VDA+LINP DS AI TA++SAN +PV+ +

Sbjct: 65 KRGMKVIIIVDAQNDSSKQTSVDLIIQOGVDALLINPTDSSAISTAVESANAVGVFVVTI 124

35

Query: 134 DRGSEGGKVLTTVASDNVAAGKMAADYAVKIKLGKAKAFELSGVPGASATVDRGKGFHSV 193

DR +E GKV T VASDNV G+MAA + KLGK AK EL GVPASAT +RG GFH++

Sbjct: 125 DRSAEQGKVETLVASDNVKGEMAAAFIADKLGKAKVAELEGVPGASATRERGSGGFHNI 184

40

Query: 194 AKSKLDILSSQSANFDRAKALNTTQNMIIQGHKDVQIIIFAQNDMALGAAQAVKSAGLQNV 253

A KL +++ QSA+FDR K L +N++QGH D+Q +FA NDEMALGA +A+ S+G +++

Sbjct: 185 ADQKLQVVTKQSAFDRITKGLTVMENLLQGHPTIQAQVFAHNDEMALGALEAINSSG-KDI 243

45

Query: 254 LIVGIDGQPDADAIKKGDISATIAQPAKMGETIAIQAAIDYYKGGKVEKETISPIYLVTK 314

L++G DG DA +IK +SAT+AQQP +G++A +AA D GKKV+K +P+ L T+

Sbjct: 244 LVIGFDGNKDALASIKDRKLSATVAQQPELIGKLATEADDILHGKVKVQKTISAPLKLETQ 304

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6478 (GBS203) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 12; MW 36.8kDa).

GBS203-His was purified as shown in Figure 208, lane 8.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2364-

Example 2093

A DNA sequence (GBSx2208) was identified in *S.agalactiae* <SEQ ID 6479> which encodes the amino acid sequence <SEQ ID 6480>. This protein is predicted to be galactoside ABC transporter, permease protein (rbsC). Analysis of this protein sequence reveals the following:

```

5   Possible site: 14
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -11.15    Transmembrane    63 - 79 ( 52 - 85)
      INTEGRAL    Likelihood = -3.66     Transmembrane    111 - 127 ( 110 - 128)
      INTEGRAL    Likelihood = -2.71     Transmembrane    168 - 184 ( 168 - 188)
10  INTEGRAL    Likelihood = -2.44     Transmembrane    189 - 205 ( 188 - 205)
      INTEGRAL    Likelihood = -0.80     Transmembrane    17 - 33 ( 17 - 33)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
15  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9287> which encodes amino acid sequence <SEQ ID 9288> was also identified.

20 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15612 GB:Z99122 ribose ABC transporter (permease) [Bacillus subtilis]
Identities = 144/211 (68%), Positives = 182/211 (86%), Gaps = 1/211 (0%)

Query: 1  MGMLNGLFISYGKLPFFIVTLATMTIFRGATLVYSNGNPITAGLSDSFLFQFLGQGYIVG 60
25  +GM+NGL I+ GK+APFI TLATMT+FRG TLVY++GNPIT GL ++ FQ G+GY +G
Sbjct: 113 LGMINGLLITKGMAPFIATLATMTVFRGLTLVYTDGNPIT-GLGTNYGFQMFGRGYFLG 171

Query: 61  IPFPVILMFLITFIILYILLHKTAFGKSVYALGGNEKAAYISGIKLNKVKIIITYTISGIMA 120
IP P I M L F+IL++LLHKT FG+ YA+GCNEKAA ISGIK+ +VK++IY+++G+++
30  Sbjct: 172 IPVPAITMVLAFFVILWVLLHKTFFGRRTYAIGGNEKAALISGIKVRVKVMIYSLAGLLS 231

Query: 121 SISGLIITSRLSSAQPTAGASYEMDAIAAVVLGGTSLSGGKGRIIGTLIGALIIGVLNNG 180
+++G I+TSRL SAQPTAG SYE+DAIAAVVLGGTSLSGG+GRI+GTLIG LIIG LNNG
35  Sbjct: 232 ALAGAILTSRLHSAQPTAGESYELDAIAAVVLGGTSLSGGRGRIVGTLIGVLIIGTLNNG 291

Query: 181 LNIIGVSAFWQQVVKGVILMAVLLDRFKVA 211
LN++GVS+F+Q VVKGVIL+AVLLDR K A
Sbjct: 292 LNLIGVSSFYQLVVKGVILIAVLLDRKKA 322

```

40 A related GBS gene <SEQ ID 8977> and protein <SEQ ID 8978> were also identified.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2094

45 A DNA sequence (GBSx2209) was identified in *S.agalactiae* <SEQ ID 6481> which encodes the amino acid sequence <SEQ ID 6482>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.12    Transmembrane    75 - 91 ( 74 - 91)
      INTEGRAL    Likelihood = -0.64    Transmembrane    96 - 112 ( 96 - 112)
50

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2365-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2095

A DNA sequence (GBSx2210) was identified in *S.agalactiae* <SEQ ID 6483> which encodes the amino acid sequence <SEQ ID 6484>. This protein is predicted to be ribose transport ATP-binding protein rbsA (rbsA). Analysis of this protein sequence reveals the following:

```

Possible site: 35
10  >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.00    Transmembrane  401 - 417 ( 401 - 417)

      ----- Final Results -----
15      bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20  >GP:CAB15611 GB:Z99122 ribose ABC transporter (ATP-binding protein)
      [Bacillus subtilis]
      Identities = 297/493 (60%), Positives = 375/493 (75%), Gaps = 1/493 (0%)

Query: 1  MKIDMRNISKSFSGTNKVLEKIDLELQSGQIHALMGENGAGKSTLMNILTGLFPASTGTIY 60
25  Sbjct: 1  MQIEMKDIHKTFGKNQVLSGVSPQLMPGEVHALMGENGAGKSTLMNILTGLHKADKGQIS 60

Query: 61  IDGEERTFSNPQEAEEFGISFIHQEMNTWPMTVLENLFLGREIKTTFGLLNQKLMRQKA 120
30  Sbjct: 61  INGNETYFSPNPKEAEQHGIAFIHQELNIWPMTVLENLFIGKEISSKLGLVLTQTRMKALA 120

Query: 121 LETFKRLGVTIPLDIPIGNLSVGQQQMIEIAKSLNQLSILVMDEPTAALTDRATENLFR 180
35  Sbjct: 121 KEQFDKLSVSLSLDQEAGECSVGQQQMIEIAKALMTNAEVIIMDEPTAALTREISKLF 180

Query: 181 VIRGLKQEGVGVVYISHRMEEIFKITDFVTVMRDGVIVDTKETSLSNDELVKKMVGRKL 240
40  Sbjct: 181 VITALKKNQVSIVYISHRMEEIFAICDRITIMRDGKTVDTTINSETDFDEVVKKMVGREL 240

Query: 241 EDYYPKHKHSEIGPVAFEVSNL-CGDNFEDVSFYVRKGEILGFSGLMGAGRTEVMRTIFGI 299
45  Sbjct: 241 TERYPKRTPSLGDKVFEVKNASVKGSEFEDVSFYVRSGEIVGVSGLMGAGRTEMMRALFGV 300

Query: 300 DKKKSGKVKIDDQEITITTPSQAIKQGIGFLTENRKDEGLILDENIKDNMTLPSTKDFSK 359
50  Sbjct: 301 DRLDTGEIWIAGKKTAKNPQEAUVKGLGFITENRKDEGLLLDTSIRENIALPNLSSFS 360

Query: 360 HGFFDEKTSTTFVQQLINRLYIKSGRPDLEVCNLSGGNQKQVVLAKWIGIAPKVLILDEP 419
55  Sbjct: 361 KGLIDHKREAEFVDLLIKRLTIKTASPTHARHLSGGNQKQVVIKWIIGIGPKVLILDEP 420

Query: 420 TRGVDVGAKREIYQLMNEADRGVPVIMVSSDLPEILGVSDRIMVMHEGRISGELSRKEA 479
      TRGVDVGAKREIY LMNEL +RGV I+MVSS+LPEILG+SDRI+V+HEGRISGE+ +EA
      Sbjct: 421 TRGVDVGAKREIYTLMNEITERGVAIIMVSSSELPEILGMSDRIIVVHEGRISGEIHARE 480

Query: 480 DQEKVMQLATGGK 492
      QE++M LATGG+
      Sbjct: 481 TQERIMTLATGGR 493

```

There is also homology to SEQ ID 4678.

-2366-

SEQ ID 6484 (GBS407d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 2-4; MW 72kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 5 & 6; MW 47kDa).

GBS407d-His was purified as shown in Figure 235, lane 9-10.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2096

- A DNA sequence (GBSx2211) was identified in *S.agalactiae* <SEQ ID 6485> which encodes the amino acid sequence <SEQ ID 6486>. This protein is predicted to be high affinity ribose transport protein rbsd (rbsD). Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2673 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:CAB15610 GB:Z99122 ribose ABC transporter (membrane protein)
[Bacillus subtilis]
Identities = 74/131 (56%), Positives = 95/131 (72%), Gaps = 1/131 (0%)

Query: 1 MKKTGILNSHLAKLADDLGHTDRVCIGDLGLFVNPNGIPKIDLSLTSGIPSFQEVLDIYLE 60
25 MKK GILNSHLAK+ DLGHTD++ I D GLPVP+G+ KIDLSL G+P+FQ+ + E
Sbjct: 1 MKKHGILNSHLAKILADLGHDTKIVADAGLPVDPGVKIDLSLKPLPAFQDTAAVLAE 60

Query: 61 NILVEKVI LAEEI KEANPDQLSRLLAKLDNSVSIEYVSHNHLKQMTQDVKAVIRTGENTP 120
+ VEKVI A EIK +N + ++ L L + IEY+SH K +T+D KAVIRTGE TP
30 Sbjct: 61 EMAVEKVI AAEI KASNQEN-AKFLNLFSEQEIEYLSHEEFKLLTKDAKAVIRTGEFTP 119

Query: 121 YSNII LQSGVI 131
Y+N ILQ+GV+
35 Sbjct: 120 YANCILQAGVL 130

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2097

- 40 A DNA sequence (GBSx2212) was identified in *S.agalactiae* <SEQ ID 6487> which encodes the amino acid sequence <SEQ ID 6488>. This protein is predicted to be ribokinase (rbsK). Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have an uncleavable N-term signal seq

- 45 ----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50

-2367-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15609 GB:Z99122 ribokinase [Bacillus subtilis]
Identities = 132/293 (45%), Positives = 177/293 (60%), Gaps = 4/293 (1%)

5   Query: 1   MSNIVIIGSISMDLVMETNRIAKEGETVFGQRFMSVPGGKGANQAVAIGRLSQERDNITI 60
      M NI +IGS SMDLV+ +++ K GETV G F VPGGKGANQAVA RL + + +
      Sbjct: 1   MRNICVIGSCSMDLVVTS DKRPKAGETVLGTSFQTVPGGKGANQAVAAARLGAQ---VFM 57

10  Query: 61   LGAIGEDSFGPILLNLNKNHVTTDFVGTIP-SSSGVAQITLYNNDNRIIYCPGANGKVD 119
      +G +G+D +G +L+NL N V TD++ + + SG A I L DN I+ GAN +
      Sbjct: 58   VGKVGDDHYGTAILNNLKANGVR TDYMEPVTHTESGTAHIVLAEGDNSIVVVGANDDIT 117

      Query: 120  TKKWSQEWSIIKEADLVVLQNEIPHQANMKIANFCKEHSIKLLYNPAPSRETDIEMLDKV 179
      I++ D+V++Q EIP + ++ +C H I ++ NPAP+R E +D
15  Sbjct: 118   PAYALNALEQIEKVDMLVLIQQEIPETVDEVCKYCHSHDIP IILNPAPARPLKQETIDHA 177

      Query: 180  DYFTPNEHECQELFPNQKLEDILATYPEKLIVTLGKGATYSDGKESHLIPALETKAVD 239
      Y TPNEHE LFP + + LA YP KL +T G +G YS G + LIP+ + VDT
20  Sbjct: 178   TYLTPNEHEASILFPBELTISEALALYPAKLFITEGKQGVRYAGSKEVLIPFPVEPVDT 237

      Query: 240  TGAGDTTFNGAFGYAISKKFKIAKALRFATLAHLVQKFGAQGGMPTTIKEMED 292
      TGAGDTFN AF A+++ I ALRFA AA LSV FGAQGGMPT E+E+
25  Sbjct: 238   TGAGDTFNAAFAVALAEGKDIEAALRFANRAASLSVCSFGAQGGMPTRNEVEE 290
```

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2098

30 A DNA sequence (GBSx2213) was identified in *S.agalactiae* <SEQ ID 6489> which encodes the amino acid sequence <SEQ ID 6490>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35      bacterial cytoplasm --- Certainty=0.2272 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

40 A related GBS nucleic acid sequence <SEQ ID 9477> which encodes amino acid sequence <SEQ ID 9478> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15608 GB:Z99122 transcriptional regulator (LacI family)
[Bacillus subtilis]
Identities = 141/327 (43%), Positives = 204/327 (62%), Gaps = 4/327 (1%)

45  Query: 13   MSTIRQVAEKAGVSTSTVSRYSISQNGYVSQKASQKIEQAIRELHYVPNFLAQSLKTKKNQ 72
      M+TI+ VA AGVS +TVSR ++ NGYV ++ ++ A+ +L+Y PN +A+SL +++++
      Sbjct: 1   MATIKDVAGAAGVSVATVSRNLNDNGYVHEETRTRVIAAMAKLNNYPNEVARSLYKRESR 60

50  Query: 73   LVGLLLPDISNPFFPRLARGVEEFLKEQGYRVMLGNTNNKSHLEEEYLNVLQSNAAAGII 132
      L+GLLLPDI+NPFFP+LARG E+ L +GYR++ GN++ + E EYL Q++ AGII
      Sbjct: 61   LIGLLLPDITNFFPQLARGAEDELNREGYRLIFGNSDEELKKELEYLQTFKQNHVAGII 120

      Query: 133  --TTHDFTKNHPEIDIPVVVVDVRVNQETQYGVFSDNKEGGKLAQAQAIWTAGATNILLIRG 190
      T + + + ++ PVV +DR E V SD G KLAQAQAI + I L+RG
55  Sbjct: 121   AATNYPDLEEYSGMNYPVVFLDR-TLEGAPSVSSDGYTGKLAQAQAIHGSQRITLLRG 179

      Query: 191   PLDKADNLNQRFGSQNYLLNKGACFAIEDSASFDAEIQTIEAKTLLDHHPDIDSIIAPS 250
```

-2368-

P RF G+ L F + ++ASF + Q AK L +P D +IA +
 Sbjct: 180 PA-HLPTAQDRFNGALEILKQAEVDFQVIETASFISIKDAQSMAKELFASYPATDGVIAN 238
 Query: 251 DIHAIAYLHEILNRGKRIPEDVQIIIGYDDILMSQFIYPSLSTIHQSSYIMGQKAAELIFK 310
 5 DI A A LHE L RGK +PED+QIIIGYDDI S ++P LSTI Q +Y MG++AA+L+
 Sbjct: 239 DIQAAAVLHEALRRGKNVPEDIQIIIGYDDIPQSGLLFPPLSTIKQPAYDMGKEAAKLLLG 298
 Query: 311 ITNQLPITNKRIKLPVHYVERETLRK 337
 I + P+ I++PV Y+ R+T R++
 10 Sbjct: 299 IIKKQPLAETAIQMPVTYIGRKTTRKE 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6491> which encodes the amino acid sequence <SEQ ID 6492>. Analysis of this protein sequence reveals the following:

Possible site: 35
 15 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1657(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 232/328 (70%), Positives = 274/328 (82%)

25 Query: 10 GVSMSTIRQVABKAGVSTSTVSRYISQNGYVSQKASQKIEQAIRELHYVPNFLAQSLKTK 69
 G +M TI+QVAB+AGVS STVSRYISQ GYVS A KI+ AI +LHY PN LAQSLKTK
 Sbjct: 14 GKAMVTIKQVABEEAGVSRSTVSRYISQKGYVSDARHKIKAAIAKLHYTPNVLAQSLKTK 73
 Query: 70 KNQLVGLLLPDISNPFPPRLARGVEEFLKEQGYRVM LGNTNNKSHLEEEYLNVLQSNAA 129
 30 KNQLVGLLLPDISNPFPPRLARG EE+LKE+GYRVM LGN ++ LEEBY++VLLQSNAA
 Sbjct: 74 KNQLVGLLLPDISNPFPPRLARGAEYLYLKEGYRVM LGNISDSEALEEEYVHVLLQSNAA 133
 Query: 130 GIITTHDFTKNHPEIDIPVVVVDRVNQETQYGVFSDNKEGKLAQAIIWTAGATNILLIR 189
 GIITTHDFTK +P + IPVVVVDRV+QETQYGVFSDN+ GG LAAQ +W AGA +LLIR
 35 Sbjct: 134 GIITTHDFTKRYPTLAIPVVVVDRVDQETQYGVFSDNRAGGLLAAQTVWQAGAKEVLLIR 193
 Query: 190 GPLDKADNLNQRFGGSQNYLLNKGACFAIEDSASFDAEIQIEAKTLLDHPDIDSIIAP 249
 GPLD A+N+N+RF+ S +YL + + DS +FDF IQ+EA L +P IDSIIAP
 40 Sbjct: 194 GPLDNAENINERFEASFSYLQKQDVTMYVCDSQNFDFESIQLSEASYNLKCYPITIDSIIAP 253
 Query: 250 SDIHAIAYLHEILNRGKRIPEDVQIIIGYDDILMSQFIYPSLSTIHQSSYIMGQKAAELIF 309
 SDIHAIAIY+HE+ ++GK+IP+DVQIIIGYDDILMSQFIYPSLSTIHQSSY+MG+ AAEL++
 Sbjct: 254 SDIHAIAIYIHELHSQGGKIPQDVQIIIGYDDILMSQFIYPSLSTIHQSSYLMGRYAAELVY 313
 Query: 310 KITNQLPITNKRIKLPVHYVERETLRK 337
 I +QL + RIKLPVHYVERET+R++
 45 Sbjct: 314 TIASQLTVKANRIKLPVHYVERETIRKR 341

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 2099

A DNA sequence (GBSx2214) was identified in *S.agalactiae* <SEQ ID 6493> which encodes the amino acid sequence <SEQ ID 6494>. Analysis of this protein sequence reveals the following:

Possible site: 57
 55 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -13.80 Transmembrane 27 - 43 (24 - 51)
 INTEGRAL Likelihood = -10.61 Transmembrane 337 - 353 (329 - 362)
 INTEGRAL Likelihood = -9.18 Transmembrane 257 - 273 (249 - 276)
 60 INTEGRAL Likelihood = -8.92 Transmembrane 302 - 318 (291 - 326)

-2369-

----- Final Results -----

bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 8979> which encodes amino acid sequence <SEQ ID 8980> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 6

SRCFLG: 0

10

McG: Length of UR: 4

Peak Value of UR: 3.20

Net Charge of CR: 1

McG: Discrim Score: 6.06

GvH: Signal Score (-7.5): 0.0500002

15

Possible site: 46

>>> Seems to have a cleavable N-term signal seq.

Amino Acid Composition: calculated from 47

ALOM program count: 3 value: -10.61 threshold: 0.0

20

INTEGRAL Likelihood = -10.61 Transmembrane 326 - 342 (318 - 348)

INTEGRAL Likelihood = -9.18 Transmembrane 246 - 262 (238 - 265)

INTEGRAL Likelihood = -8.92 Transmembrane 291 - 307 (280 - 315)

PERIPHERAL Likelihood = 4.98 152

modified ALOM score: 2.62

icml HYPID: 7 CFP: 0.525

25

*** Reasoning Step: 3

----- Final Results -----

30

bacterial membrane --- Certainty=0.5246(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35

>GP:AAF12525 GB:AE001863 hypothetical protein [Deinococcus radiodurans]
 Identities = 103/352 (29%), Positives = 191/352 (54%), Gaps = 9/352 (2%)

Query: 15 AWKELTFYKKYLLIELLIIVMMFMVFLSGLANGLGRAVSAAIENNPQTYILNEGAEQ 74

A +EL K + LLI ++ ++ FMV L+GL GL R ++ + + PAQ+++ + A+

40

Sbjct: 4 ALRELQHQKLRSLLIGGIVALIAFMVFMVLTGLTRGLSRDSASLLLDTPAQSFVITKEADG 63

Query: 75 VITSSVLITKQTDNLNLSNLKDSITLNIQRSSLTRQGEKKIDISYFAIDKDSFMAPTLS 134

V+ S L+ + +++L + ++ ++ +K++ +D F+AP +S

Sbjct: 64 VLNRSFLSPEQ--VSALQQDNEDAAFAQTFSFSGDKQLSGVLLGVDPRGFLAPDVS 120

45

Query: 135 EGKQLTSYKKAIIILNDSLKAEGIKLGDKVIDKSSSISLTVVGFVHNSMYGHGPVAFIDKD 194

EG+ L A++ ++SL+ +G+K+GD + K S L V GF ++ H P ++

Sbjct: 121 EGQTLRVAGGAVV-DESLREDGVKVGDLTLKPSCDQLRVSGFTRSARLNHQPGMYVSLA 179

50

Query: 195 IYTEINKKINPQYQFLPQALVMKNDKSISHLP-TQLEAVSKKDVIQHHPGYSAEQSTLNM 253

+ +K+NP+ A+ + + +L L ++ +Q +PGY EQ +L M

Sbjct: 180 RW----QKLNPRMHGTVNAVALPAAPQVNLGGADLSVTNRAQTLQVLPQYKEEQGSLTM 235

55

Query: 254 ILWVLVVASAGILGVFFYIITLQKRHEFSVMKAIGTKMSEIALFQLSQVILALFGIIVG 313

I L+ +A +L FFY++TLQK +F ++KAIG +A ++Q++IL L + +

Sbjct: 236 IQVFLIAVAAFVLATFFYVMTLQKTAQFGLLKAIGASNRTLGASVVAQMLILTLLAVAIA 295

60

Query: 314 DGLAVALSYVLPAQMPFVINWQNIILVSFVFLVIAMISSALSIVKAKIDPV 365

+ + + +LPA MPF + NI S + LV+A ++S LS+ +VAK+DP+

Sbjct: 296 AAVTLGMVQLLPAGMPFHLTAANIASASGLLLVVAALASLLSVRRVAKVDPL 347

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6495> which encodes the amino acid sequence <SEQ ID 6496>. Analysis of this protein sequence reveals the following:

Possible site: 58

-2370-

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -12.31 Transmembrane 246 - 262 (233 - 270)
 INTEGRAL Likelihood = -8.49 Transmembrane 327 - 343 (321 - 351)
 INTEGRAL Likelihood = -1.01 Transmembrane 301 - 317 (301 - 317)

----- Final Results -----

bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF12525 GB:AE001863 hypothetical protein [Deinococcus radiodurans]
 Identities = 101/360 (28%), Positives = 175/360 (48%), Gaps = 11/360 (3%)

Query: 1 MFLALNEMKQSKLRYGLIAGLLCLVAYLMFFLSGLAFGLMQENRSAVDLWKADSVLLAKD 60
 M+LAL E++ KLR LI G++ L+A+++F L+GL GL +++ S + A S + K+
 Sbjct: 1 MYLALRELQHQKLRSLIGGIVALIAFMVFMILTGLTRGLSRDSASLLLDTPAQSFVTTKE 60

Query: 61 ADATLTLQSQVSRAQENQITADKVAPLAQLNTVAWSVKNPKDADKVKVSLFGIDSNSFIRP 120
 AD L S +S Q + + D A T K V L G+D F+ P
 Sbjct: 61 ADGVLNRSFLSPEQVSALQQDNEDAAFAQTFVSFSGDKQLSGV---LLGVDPGRGFLAP 117

Query: 121 NIVKGRLFKTNKEVVLDQSLAKEEFAIGKDFYTSSSSQALTIVGYTQNFARFSVAPVVVM 180
 ++ +G+ + V+D+SL +E+ +G S L + G+T++AR + P +Y+
 Sbjct: 118 DVSEGGTLRVAGGAVVDESL-REDGVKVGDLVTLKPSGDQLRVSGFTRARLNHQPGMYV 176

Query: 181 NLEAFETLKYGEPLPKDKQVVNAFITKGS--LTDYPKDFQKLDIKFTITKLPGYSAQLL 238
 +L ++ L P+ VNA + + D + + LPGY +
 Sbjct: 177 SLARWQKLN-----PRMHGTVNAVALPAAPQVNLGGADLSVTNRAQTLQVLPGYKEEQG 231

Query: 239 TFGFMISFLVIISAIIGIFMYILTIQKAPIFGIMKAQGISNKTITTAFLMQTFFLSFLG 298
 + + FL+ ++A ++ F Y++T+QK FG++KA G SN+T+ +V+ Q L+ L
 Sbjct: 232 SLTMIQVFLIAVAFLATFFVYMTLQKTAQFGLLKAIGASNRTLAGSVVAQMLILTLA 291

Query: 299 SGLGLLGTWLTSLLLPTVVPFQSNWFLYLAIFVSMICFALLGTLFSVFNIIRIDPLKAIG 358
 + T LLP +PF + ++ A L +L SV + ++DPL A+G
 Sbjct: 292 VAIAAAVTLGMVQLLPAGMPFHLTAANIASASGLLLVVAALASLLSVRRVAKVDPLIALG 351

An alignment of the GAS and GBS proteins is shown below.

Identities = 96/356 (26%), Positives = 178/356 (49%), Gaps = 4/356 (1%)

Query: 15 AWKELTFYKKYLLIELLIIVMMFMVFLSGLANGLGRAVSAAIENNPACTYILNEGAEQ 74
 A E+ K +Y LI L+ ++ +++ FL SGLA GL + +A++ A + +L + A+
 Sbjct: 4 ALNEMKQSKLRYGLIAGLLCLVAYLMFFLSGLAFGLMQENRSAVDLWKADSVLLAKDADA 63

Query: 75 VITSSVLTTKDQTDLNSLNLDKSTTLNIQRSSSLTRQGHEKKIDISYFAIDKDSFMAPTLS 134
 +T S ++ + + + + LN S+ K+ +S F ID +SF+ P +
 Sbjct: 64 TLTLQSQVSRAQENQITADKVAPLAQLNTVAWSVKNPKDADKVKVSLFGIDSNSFIRPNIV 123

Query: 135 EGKQLTSYKKAIIILNDSLKAEGIKLGDKVIDKSSSISLTVVGFVHNSMYGHGPVAFIDKD 194
 +G+ + K+ ++ K E +G SSS +LT+VG+ N+ + PV +++ +
 Sbjct: 124 KGRLEKTNKEVVLDQSLAKEEFAIGKDFYTSSSSQALTIVGYTQNFARFSVAPVVVMNLE 183

Query: 195 IYTEIN-KKINPQVQFLPQALVMKNDKSIHSLPTQ-LEAVSKKDVIQHIPGYSABEQSTLN 252
 + + + P+ ++ A + K S++ P + + + K I +PGYSA+ T
 Sbjct: 184 AFETLKYGEPLPKDKQVVNAFITKG--SLTDYPKDFQKLDIKFTITKLPGYSAQLLTFG 241

Query: 253 MILWVLVVASAGILGVFFYIITLQKRHEFSVMKAIGTKMSEIALFQLSQVILALFGIIV 312
 ++ LV+ SA I+G+F YI+T+QK F +MKA G I L Q L+ G +
 Sbjct: 242 FMISFLVIISAIIGIFMYILTIQKAPIFGIMKAQGISNKTITTAFLMQTFFLSFLGSG 301

Query: 313 GDGLAVALSYVLPAQMPFVINWQNIILVSFVFLVIAMISSALSIVKAKIDPVEVI 368
 G S +LP +PF NW + + + A++ + S+ + +IDP++ I
 Sbjct: 302 GLLGTLWLTSLLLPTVVPFQSNWFLYLAIFVSMICFALLGTLFSVFNIIRIDPLKAI 357

-2371-

SEQ ID 8980 (GBS239) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 13; MW 64kDa).

GBS239-GST was purified as shown in Figure 227, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2100

A DNA sequence (GBSx2215) was identified in *S.agalactiae* <SEQ ID 6497> which encodes the amino acid sequence <SEQ ID 6498>. This protein is predicted to be heterocyst maturation protein (devA) (b0879). Analysis of this protein sequence reveals the following:

```

10   Possible site: 33
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
15   bacterial cytoplasm --- Certainty=0.1751(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20   >GP:CAA05977 GB:AJ003195 ATP-binding subunit [Anabaena variabilis]
    Identities = 87/225 (38%), Positives = 146/225 (64%), Gaps = 1/225 (0%)

    Query: 3  AILELKHISKHYPDGDELLSILDNLDSVSAGEFVAILGPSGSGKSTLLSIAGLLLGADQ 62
              A++ +K ++ +Y G      IL +++L +  GE V + GPSGSGK+TLLS+ G L      +
25   Sbjct: 5  AVIAIKSLNHYYGKALKRQILFDINLEIYPGEIVIMTGPSGSGKTTLLSLIGGLRSVQE 64

    Query: 63  GSLYVNHENVTDLRSQRQRTQLRREALGFIFQSHQLLPYLTIQEQLEAFKHYDKKTS 122
              G+L      ++ SQ +  Q+RR ++G+IFQ+H LL +LT ++ +Q      +H ++ +
30   Sbjct: 65  GNLQFLGVELSGASQNKLVQIRR-SIGYIFQAHNLLGFLTARQNVQMAVELNEHISQEEA 123

    Query: 123 LEEINKLLSDLGIEQCAHKYPNQLSGGQKQRAAIARAFINHPKVLADEPTASLDEERGR 182
              + +  +L  +G+E      YP+ LSGGQKQR AIARA +N+P ++LADEPTA+LD++ GR
35   Sbjct: 124 IAKAEAMLKAVGLENRVDDYYPDNLSGGQKQRVAIARALVNNPPLVLADEPTAALDKQSGR 183

    Query: 183 QVTELIRQEVKSHNTAAIMVTHDERVLDLVDTVYRLKDGKLVKEN 227
              V E++++ K  T+ ++VTHD R+LD+ D +  ++DG L +++
40   Sbjct: 184 DVVEIMQRLAKDQGTSILLVTHDNRIIDIADRIVEMEDGILARDS 228

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6499> which encodes the amino acid sequence <SEQ ID 6500>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 13
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
45   bacterial cytoplasm --- Certainty=0.4181(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

50   Identities = 103/224 (45%), Positives = 149/224 (65%), Gaps = 4/224 (1%)

    Query: 3  AILELKHISKHYPDGDELLSILDNLDSVSAGEFVAILGPSGSGKSTLLSIAGLLLGADQ 62
              ++L K ++K + DG  ++ L  D S+ AGEFVAI+GPSGSGKST L+IAG L
55   Sbjct: 3  SVLTFKQVTKTFQDGHHEINALKATDFSIEAGEFVAIIGPSGSGKSTFLTIAGGLQTPSS 62

    Query: 63  GSLYVNHENVTDLRSQRQRTQLRREALGFIFQSHQLLPYLTIQEQLEAFKHYDKKTS 122

```

-2372-

G L *++ + T Ls+++R++LR +++GFI Q+ L+P+ T+Q+QL+ H
 Sb jct: 63 GQLIIDGTDYTHLSEKERSRLRFKSVGFILQASNLIPFSTVQQLE----LVDHLTGSKE 118
 Query: 123 LEEINKLLSDLGIEQCAHKYPNQLSGGQKQRAAIARAFINHPKVILADEPTASLDEERGR 182
 + N+L DLGI H+ P +LSSG++QRAAIARA + P +ILADEPTASLD E+
 Sb jct: 119 KAKANQLFDDLGITGLKHQLPQELSGGERQRAAIARALYHDPALILADEPTASLDEKAY 178
 Query: 183 QVTLEIRQEVKSHNTAAIMVTHDERVLDLVDTVYRLKDGKLVKE 226
 +V +L+ +E K N A IMVTHD+R+L D VYR++DG+L +E
 Sb jct: 179 EVVKLLAKESKEKNKAIIMVTHDDRMLKYCDKVYRMQDGELCQE 222

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2101

A DNA sequence (GBSx2216) was identified in *S.agalactiae* <SEQ ID 6501> which encodes the amino acid sequence <SEQ ID 6502>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2645(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 The protein has homology with the following sequences in the GENPEPT database.
 >GP:CAB64972 GB:AJ012050 VicR protein [Enterococcus faecalis]
 Identities = 86/229 (37%), Positives = 132/229 (57%), Gaps = 10/229 (4%)
 Query: 3 KILVVEDNIVQQKIITTKLTQEGYQFTASNGQEALNCLDTEEVQLIITDIMPMDGYQ 62
 KILVV+D +I+ L +EGY+ TA +G+EAL ++ E LII D+M+P MDG +
 Sb jct: 52 KILVVDDEKPISEIVKYNLVKEGYEVFTAYDGEEALEKVEEVEPDLIILDLMLPKMDGLE 111
 Query: 63 LIQELRSAAYNVPIIVMTAKSQMEDMTKGFGGLGADDYMKVPVQLQELALRIKALLRR--- 119
 + +E+R +++PII++TAK D G LGADDY+ KP +EL R+KA LRR
 Sb jct: 112 VAREVRK-THDMFIIMVTAKDSEIDKVLGLELGADDYVTKPFSNRELVARVKANLRRGAT 170
 Query: 120 ----ANIVAQHQLIIGNTCLNEDELSLKYFEQEIIFPQKEFRVLFHLLSYNRIFRLEL 175
 A + Q +L IG+ ++ D + ++I +EF +L++L + ++ TR L
 Sb jct: 171 NAKAEAVTTQSELTIGDLTIHPDAYMVSKRGEKIELTHREFELLYLAKHIGQVMTREHL 230
 Query: 176 LDSIWGMDTDLDERVVDACINKIRKVEHLPDFK--IETVRGVGYRAKN 222
 L ++WG D D R VD + ++R K+E P + T RGVGY +N
 Sb jct: 231 LQTVWGYDYFGDVRTVDVTVRRLREKIEDSPSHPTYLVTTRGVGYLRLN 279

There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2102

A DNA sequence (GBSx2217) was identified in *S.agalactiae* <SEQ ID 6503> which encodes the amino acid sequence <SEQ ID 6504>. This protein is predicted to be sensor protein. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -8.97 Transmembrane 53 - 69 (47 - 77)

-2373-

----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC62214 GB:AF049873 sensor protein [Lactococcus lactis]
 Identities = 97/307 (31%), Positives = 169/307 (54%), Gaps = 16/307 (5%)

Query: 57 SALAVVFLSLVIASISMWYGSYHLTKPILDISHIVSNVADGDFEGHIYRNSNRKSYEY 116
 + LAV+ +L++ + S++Y + +T+P+L I +A GD + N+
 Sbjct: 170 AVLAVI--TLIVTAFSIFYITRTVTRPLLKIKLGTDKIAQQDLSIQLNVNTE----- 219

Query: 117 NELDELSEINQMIVSLSHMDHMRKDFITNVSHLKTPIAAVANIVELLQDPELDEETQS 176
 +EL EL++SI + L M R +F+++V+HEL+TP+ + ++ E ++
 Sbjct: 220 DELGELAKSIEDLAELKDFMKRERNEFLSSVAHELRTPLTFIKGYADIANRSTTSLEDKT 279

Query: 177 ELILGLVKTESLRRLRLCDTMLQMSRVDNQETIGELSSVRVDEQIRQAMISLTERWQAKRI 236
 + L +++ ES LT+L + ++ +++++ E V + E I + + ++ + KRI
 Sbjct: 280 QYLRIRIREESSRHITQLMEDLMNLAQLEENGFKVEKHQVLIQELINEVSVKSGVFSEKRI 339

Query: 237 NFQLDSKPYTVVYNSNDLLM--QVWINLLDNAIKYSEDIVDLVSRMEETNNHYLRVTISDK 294
 NF L S Y+N D + QV +NLL NA KYS D D+ + ++ +++ISDK
 Sbjct: 340 NF-LISGEGNFYANIDFMRIEQVLVNLMLNAYKYSADESDIKLAFIPEKENF-KIVISDK 397

Query: 295 GRGISQYDVQHIFDKFYQADQSHNQ--GNGLGLAIVKRIIVLCKGRISVSSQLBIGTF 352
 G GI + D+ +IF++FY+ D+S + G GLGLAIV+ I+ G+I V S GT F
 Sbjct: 398 GEGIPEQDLPIYIFERFYRVDKSRTRTTGGVGLGLAIVQDIVKKNHNGKIIVESIQNGTTF 457

Query: 353 CVELPLS 359
 +ELP S
 Sbjct: 458 IIELPYS 464

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8981> and protein <SEQ ID 8982> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 4.84
 GvH: Signal Score (-7.5): 0.179999
 Possible site: 35
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 1 value: -8.97 threshold: 0.0
 INTEGRAL Likelihood = -8.97 Transmembrane 50 - 66 (47 - 77)
 PERIPHERAL Likelihood = 1.27 324
 modified ALOM score: 2.29

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

31.9/57.3% over 293aa

Lactococcus lactis

GP|3687664| sensor protein Insert characterized

ORF01881(478 - 1377 of 1677)
 GP|3687664|gb|AAC62214.1||AF049873(171 - 464 of 464) sensor protein {Lactococcus lactis}
 %Match = 12.9

-2374-

```

%Identity = 31.9 %Similarity = 57.3
Matches = 94 Mismatches = 121 Conservative Sub.s = 75

5      339      369      399      429      459      489      519      549
MTKLRFRFRFLRFYFTLMFVLTMLFSVLASLLLVAAIVFTFFQGVLTTHVLQVSALAVVFLSLVIASISMWYGSYHLTKP
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
EKKNKKESLHFHWLGDKYIVSKSRIQSNGKIVGSVYMFLLSTRPIQKMFNFNTGIFAVLAVITLIVTAFSIFYITRTVTRP
      130      140      150      160      170      180      190

10     579      609      639      669      699      729      759      789
ILDISHIVSNVADGDFEGHIYRNSNRKSYEYYNELDELSINQMIVSLSHMDHMRKDFITNVSHELKTPIAAVANIVE
:|:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
LLKIKLGTDKIAQGDLSIQLNVNTE-----DELGELAKSIEDLAEKLDPMKRENEFLSSVAHELRTPLTFIKGYAD
      210      220      230      240      250      260

15     819      849      879      909      939      969      999      1029
LLQDPELDEETQSELLGLVKTESLRLTRLCDTLMQMSRVDNQETIGELSSVRVDEQIRQAMISLTERWQAKRINFQDLSK
:|:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
IANRSTTSLEDKTQYLRRIRESRHLTQLMEDLMNLAQLEENGFKVEKHQVLIQELINEVSVKSGVGFSEKRINF-LISG
      280      290      300      310      320      330      340

20     1059     1083     1113     1143     1173     1203     1233
PYTVYSNSDLL--MQVWINLLDNAIKYSEDIVDLSVRMEETNNHYLRVVISDKGRGISQYDVQHIQADQSHNQ-
|:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
EGNFYANIDFMRIEQVLVNLLMAYKYSADESDIKLAFIPEKENF-KIVISDKGEGIPEQDLPIYIFERFYRVDKSRTRTT
      360      370      380      390      400      410      420

25     1287     1317     1347     1377     1407     1437     1467     1497
-NGNLGLAIVKRIIVLCKGRISVSSQLEIGTEFCVELPLS*LFKTTITANWQLLFYLFNRKYTKNRQKL*KYLITINIASV*
|:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GGVGLGLAIVQDIVKHNKGIIVESIQNQGTTFIIELPYS
      440      450      460

```

SEQ ID 8982 (GBS170d) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 181 (lane 4; MW 35kDa) and in Figure 123 (lane 5-7; MW 35kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 2-4; MW 60kDa) and in Figure 184 (lane 3; MW 60kDa). Purified GBS170d-GST is shown in Figure 243, lane 7; purified GBS170d-His is shown in Figure 234, lanes 5-6.

Example 2103

A DNA sequence (GBSx2218) was identified in *S. agalactiae* <SEQ ID 6505> which encodes the amino acid sequence <SEQ ID 6506>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0502 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50 >GP:BA06906 GB:AP001518 argininosuccinate synthase
(citrulline-aspartate ligase) [Bacillus halodurans]
Identities = 262/396 (66%), Positives = 321/396 (80%), Gaps = 1/396 (0%)

55 Query: 1 MGKEKLILAYSGGLDTSVAIAWLK-KDYDVIAVCMVDVGEKDLDFIHDKALTIGAIESYI 59
M K+K++LAYSGGLDTSVAI WL K YDVIAV +DVGEKDL+F+ +KAL +GAIESY
Sbjct: 1 MSKKKVVLAYSGGLDTSVAIKWLSDKGYDVIAVGLDVGEKDLBFVKEKALKVGAIESYT 60

Query: 60 LDVDEFAEHFVLPALQAHAMYEQKYPLVSALS SRPIIAQKLIVEMAHQTGATTIAHGCTGK 119
+D K EF AE FVL PALQAHA +YEQKYPLVSALS SRP +I++KLVE+A QTGA +AHGCTGK
60 Sbjct: 61 IDAKKEFAEEFVLPALQAHALYEQKYPLVSALS SRPLISKLVIEAEQTGAQAVAHGCTGK 120

```


-2375-

5 Query: 120 GNDQVRFEVAIAALDPELVKVIAPVREWKWHREEEITFAKANGVPIPADLNPYSIDQNLW 179
 GNDQVRFEV+I AL+P L+V+APVREW W R+EEL +AK N +PIP DLDPNYS+DQNLW
 Sbjct: 121 GNDQVRFEVSIQALNPENLEVLAPVREWAWSRDEEIEYAKKNIPIDLDNPYSVDQNLW 180

10 Query: 180 GRANECGVLENPNQAPPEAFGITKSPBEAPDCAEYIDITFQNGKPIAINNQEMTLADLI 239
 GR+NECG+LE+PW PE A+ +T + E+APD E ++I F+ G P+ +N + + +LI
 Sbjct: 181 GRSNECGILEDPWATPPEGAYELTVAIEDAPDQPEIVEIGFEKGIPVTILNGKSYPVHELI 240

15 Query: 240 LSLNEIAGKHGIGRIDHVENRLVGIKSREIYECPAAMVLLAAHKEIEDLTLVREVSHFKP 299
 L LN+IAGKHG+GRIDHVENRLVGIKSRE+YECF AM L+ AHKE+EDLTL +EV+HFKP
 Sbjct: 241 LELNQIAGKHGCVGRIDHVENRLVGIKSREVYECPGAMTLIKAHKELEDLTLTKVAHFKP 300

20 Query: 300 ILENELSNLIYNALWFSPATKAIITAYVKETQKVNGTTKVKLYKGSQAQVVARHSSNSLYD 359
 ++E +++ LIY LWFSP A+ A++KETQ V G +VKL+KG A V R S SLY+
 Sbjct: 301 VVEKKIAELIYEGWFSPLQFALSFLKETQSTVTGVVRVKLFKGHAIVEGRKSEYSLYN 360

Query: 360 ENLATYTAADSFDQDAAVGFIKLWGLPTQVNAQVVK 395
 E LATYT D FD +AAVGFI LWGLPT+V + VNK
 Sbjct: 361 EKLATYTPDDEFDHNAAVGFISLWGLPTKVYSMVNK 396

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2104

A DNA sequence (GBSx2219) was identified in *S.agalactiae* <SEQ ID 6507> which encodes the amino acid sequence <SEQ ID 6508>. This protein is predicted to be argininosuccinate lyase (argH). Analysis of this protein sequence reveals the following:

30 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2131(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BA06905 GB:AP001518 argininosuccinate lyase [Bacillus halodurans]
 Identities = 284/454 (62%), Positives = 350/454 (76%)

Query: 6 KIWGGRFESSLEKWEFEFGASISFDQKLAPYDMKASMAHVMTLKGKTDIISQEEAGLIKDG 65
 KIWGGRF + E WV+EFGASI FDQ+L D++ S+AHVTML K+ I++ EE IK G
 Sbjct: 3 KIWGGRFTKTAEAWVDEFGASIGFDQQLVEEDIEGSLAHVTMLEKSGILANEEVEQIKKG 62

45 Query: 66 LKILQDKYRAGQLTFSISNEDIHNMIESLLTAEIGEVAGKLHTARSNDQVATDMHLYLK 125
 L IL +K + G+L +S++NEDIH+NIE LL EIG V GKLHT RSRNDQVATDMHLYL+
 Sbjct: 63 LHILLEKAKKGELNYSVANEDIHNLIEKLLIDEIGPVGGKLHTGRSRNDQVATDMHLYLR 122

50 Query: 126 DKLQEMMKLLHLRRTTLVNLAENHIYTVMPGYTHLQHAQPISEFGHLMAYYNMFTRDTER 185
 + +E+++ + +++ LV A+ H+ T++PGYTHLQ AQPISE HHL+AY+ M RD R
 Sbjct: 123 KQTKEILQLVKNVQAALVEQAKQHVTETLIPGYTHLQRAQPISEFAHLLAYFWMLERDYGR 182

55 Query: 186 LEFNMKHTNLSPLGAAALAGTTFFPIDRHMTTLLDDEFKPYNSLDAVSDRDFIIEFLSNA 245
 E ++K N+SPLGA ALAGTTFFPIDR T LL F+ Y NSLDAVSDRDFI+EFLS +
 Sbjct: 183 YEDSLKRLNVSPLGAGALAGTTFFPIDREYTAELLGFDGIYENSLDAVSDRDFIVEFLSAS 242

60 Query: 246 SILMMHLSRFCEEIINWCSVEYQFITLSDTFSTGSSIMPQKKNPDMAELIRGKTGRVYGN 305
 S+LM HLSR CEE+I W S E+QF+ + D F+TGSSIMPQKKNPDMAELIRGKTGRVYG+
 Sbjct: 243 SLLMTHLSRLCBELIILWSSQEFQFVEMDDAFATGSSIMPQKKNPDMAELIRGKTGRVYGS 302

-2376-

Query: 306 LFSLLTVMKSLPLAYNKDLQEDKEGMFDSVETVSIATIEIMANMLETMTVNEHIMMTSTET 365
 LFSLLTV+K LPLAYNKD+QEDKEGMFD+V+TV ++ I A M++TM V E M +
 Sbjct: 303 LFSLLTVLKGLPLAYNKDMQEDKEGMFDVAVKTVKGS LAIFAGMIQTMKVKEETMTKAVHQ 362

5 Query: 366 DFSNATELADYLASKGVFFRKAHEIVGKLVLECSKNGSYLQDIPLKYYQEISELIENDIY 425
 DFSNATELADYLA+KG+PFR+AHE+VGKLV L C + G YL D+PL Y+ S+L + DIY
 Sbjct: 363 DFSNATELADYLATKGMFFREAEHVVGKLVLLCTQKGIYLLDLPLSDYKAASDLFDEDIY 422

10 Query: 426 EILTAKTAVKRRNSLGGTGFDQVKKQILLARKEL 459
 ++L KT V RR S GGTGF +VKK I A K L
 Sbjct: 423 DVLQPKTVVARRTSAGGTGFTEVKKAKAKAEKIL 456

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 2105

A DNA sequence (GBSx2220) was identified in *S. agalactiae* <SEQ ID 6509> which encodes the amino acid sequence <SEQ ID 6510>. This protein is predicted to be class-II aldolase (fba). Analysis of this protein sequence reveals the following:

20 Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2930(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9289> which encodes amino acid sequence <SEQ ID 9290> was also identified. Analysis of this sequence reveals:

30 GvH: Signal Score (-7.5): -2.92
 Possible site: 42
 >>> Seems to have no N-terminal signal seq.
 ALOM program count: 0 value: 0.37 threshold: 0.0
 PERIPHERAL Likelihood = 0.37 66
 35 modified ALOM score: -0.57

*** Reasoning Step: 3

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.2930(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA16889 GB:AB050113 class-II aldolase [Streptococcus bovis]
 Identities = 221/242 (91%), Positives = 234/242 (96%)

50 Query: 1 MAIVSAEKFVQAARDNGYAVGGFNTNNLEWTQAILRAAEAKKAPVLIQTSMGAAKYMGGY 60
 MAIVSAEKF++AAR+NGYAVGGFNTNNLEWTQAILRAAEAKKAP+LIQTSMGAAKYMGGY
 Sbjct: 1 MAIVSAEKFIKAARENGYAVGGFNTNNLEWTQAILRAAEAKKAPILIQTSMGAAKYMGGY 60

Query: 61 KLCKQLIETLVESMGITVPVAIHLDHGHYDDALECIEVGYSIMFDGSHLPVEENLEKAR 120
 KLCK LIE LVESMGITVPVAIHLDHGH++DALECIEVGYS+MFDGSHLPVEENLEKA+
 55 Sbjct: 61 KLCKTLIENLVESMGITVPVAIHLDHGHFEDALECIEVGYSVMFDGSHLPVEENLEKAK 120

Query: 121 EVVAKAHAKGISVEAEVGTIGGEEDGIVGKGELAPIEDAKAMVETGIDFLAAGIGNIHGP 180
 EVVAKAHAKG+SVEAEVGTIGGEEDGIVG GELAPIEDAKAMV TGIDFLAAGIGNIHGP
 Sbjct: 121 EVVAKAHAKGVSVEAEVGTIGGEEDGIVGGELAPIEDAKAMVATGIDFLAAGIGNIHGP 180

-2377-

Query: 181 YPANWEGLDLHDLKLTAVPGFPFIVLHGGSGIPDDQIQEAIKLGVAKVNVTQCQLAFC 240
 YPANW+GL LDHLKLT AVPGFPFIVLHGGSGIPDDQI+ AIKLGVAKVNVTQCQ+AF
 Sbjct: 181 YPANWQGLHDLHDLKLTAAVPGFPFIVLHGGSGIPDDQIKAAIKLGVAKVNVTQCQIAFA 240

Query: 241 QA 242
 +A
 Sbjct: 241 KA 242

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6511> which encodes the amino acid sequence <SEQ ID 6512>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2930(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 217/242 (89%), Positives = 228/242 (93%)

Query: 1 MAIVSAEKFFVQAARDNGYAVGGFNTNNLEWTQAILRAAEAKKAPVLIQTSMGAAKYMGGY 60
 MAIVSAEKFFVQAAR+NGYAVGGFNTNNLEWTQAILRAAEAK+APVLIQTSMGAAKYMGGY
 Sbjct: 1 MAIVSAEKFFVQAARENGYAVGGFNTNNLEWTQAILRAAEAKQAPVLIQTSMGAAKYMGGY 60

Query: 61 KLCKQLIETLVESMGITVPVAIHLDDHGHYDDALECIEVGYSIMFDGSHLPVEENLEKAR 120
 K+C+ LI LVESMGITVPVAIHLDDHGHY+DALECIEVGYSIMFDGSHLPVEENL K
 Sbjct: 61 KVCQSLITNLVESMGITVPVAIHLDDHGHYEDALECIEVGYSIMFDGSHLPVEENLAKTA 120

Query: 121 EVVAKAHAKGISVEAEVGTIGGEEDGIVGKGLAPIEDAKAMVETGIDFLAAGIGNIHGP 180
 EVV AIAKG+SVEAEVGTIGGEEDGI+GKGLAPIEDAKAMVETGIDFLAAGIGNIHGP
 Sbjct: 121 EVVKIAHAKGVSVEAEVGTIGGEEDGIIGKGLAPIEDAKAMVETGIDFLAAGIGNIHGP 180

Query: 181 YPANWEGLDLHDLKLTAVPGFPFIVLHGGSGIPDDQIQEAIKLGVAKVNVTQCQLAFC 240
 YP NWEGL LDHL+KLT AVPGFPFIVLHGGSGIPDDQI+EAI+LGVAKVNVT Q+AF
 Sbjct: 181 YPENWEGALDHLKLTAAVPGFPFIVLHGGSGIPDDQIKEAIRLGVAKVNVTESQIAFS 240

Query: 241 QA 242
 A
 Sbjct: 241 NA 242

SEQ ID 9290 (GBS683) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 8 & 10; MW 55kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 11-13; MW 30kDa) and in Figure 184 (lane 11; MW 30kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2106

- 50 A DNA sequence (GBSx2221) was identified in *S.galactiae* <SEQ ID 6513> which encodes the amino acid sequence <SEQ ID 6514>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2775(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2378-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAA88585 GB:M18954 unknown protein [Streptococcus mutans]
 Identities = 109/229 (47%), Positives = 156/229 (67%), Gaps = 1/229 (0%)

Query: 1 MFSGKRLKKRRITLGYSQSELADKLHINRSSFYNWENKTKPNQSNLKQLAILLDVPETY 60
 MFS ++LK+RR LG SQ++ ADKL I+R SYFNWE KTKPNQ NL +LA LL V Y
 10 Sbjct: 1 MFSSQKLKERRKKLGLSQAQTADKLGISRPSYFNWWEIGKTKPNQKNLDKLAHLLKVDSAY 60

Query: 61 FESEYKIVNTYLLQLSLQNQEKVEKYAEELLQTQKVHEKIVPLFAVEVLSEIQLSAGPGEG 120
 F S++ IV Y +L+ N+ K KY++ LL+ Q ++ +LSAG G
 15 Sbjct: 61 FLSQHDIVEIYTRLNESNKTTLKYSQHLLQQDKRNLNMKNRYPYRVYEKLSAGTGYS 120

Query: 121 LYDEFETETVYSEDEYTGFDIATWISGNSMEPVYKDGFEVALIRSTGFDHGDGAVYALNWNG 180
 + + +TV+ ++E D A+WI G+SMEP++ +GEVALI+ TGFD+DGA+YA++W+G
 15 Sbjct: 121 YFGDGNFDTVFYDEEID-HDFASWIFGDSMEPIFLNGEVALIKQTGFDDYDGA+Y+DWDG 179

Query: 181 SLYIKKLYREEDGFRMVSINPDVAERFIPFEDEIRIVGKIVGHFMPVIG 229
 YIKK+YREE G R+VS+N A++F P+++ RI+G IVG+F+P+ G
 20 Sbjct: 180 QTYIKKVYREETGLRLVSLNKKYADKFAPYDENPRIIGLIVGNFIPLEG 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6515> which encodes the amino acid sequence <SEQ ID 6516>. Analysis of this protein sequence reveals the following:

25 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4340(Affirmative) < succ>
 30 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/209 (40%), Positives = 130/209 (62%), Gaps = 9/209 (4%)

35 Query: 25 LHINRSSFYNWENKTKPNQSNLKQLAILLDVPETYFESEYKIVNTYLLQLSLQNQEKVEK 84
 LH+N+ + NWE K PN+ +L L L +V YF+ Y+++ Y QL++ N+EKV
 Sbjct: 5 LHVNKMTISNWEKGNIPNEKHLNALLHLFNVTSYDFDPNVRLLTPYNQLTISNKEKVIG 64

40 Query: 85 YAEELLQTQ-----KVHEKIVPLFAVEVLSEIQLSAGPGEGLYDEFETETVYSEDEYTG 138
 Y+E LL Q + +K L+A V LSAG G + + + V+ DE
 Sbjct: 65 YSERLLNHQIDKKSKDLIDKPSQLYAYRVYES--LSAGTGYSYFGDGNFDDVVFY-DEQLE 121

45 Query: 139 FDIATWISGNSMEPVYKDGFEVALIRSTGFDHGDGAVYALNWNGSLYIKKLYREEDGFRMVS 198
 +D A+W+ G+SMEP Y +GEV LI+ FD+DGA+YA+ W+G YIKK++RE++G R+VS
 Sbjct: 122 YDFASWVFGDSMEPTYLNGEVVLIKQNSFDYDGAIVAVEWDGQTYIKKVFREDEGLRLVS 181

Query: 199 INPDVAERFIPFEDEIRIVGKIVGHFMPV 227
 +N +++F P+ +E RI+GKI+ +F P+
 50 Sbjct: 182 LNKKYSDKFAPYSEPRIIGKIIFRPL 210

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2107

55 A DNA sequence (GBSx2222) was identified in *S.agalactiae* <SEQ ID 6517> which encodes the amino acid sequence <SEQ ID 6518>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

-2379-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2387(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 **Example 2108**

A DNA sequence (GBSx2223) was identified in *S.agalactiae* <SEQ ID 6519> which encodes the amino acid sequence <SEQ ID 6520>. This protein is predicted to be UmuC MucB homolog (uvrX). Analysis of this protein sequence reveals the following:

Possible site: 47

15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2195(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9925> which encodes amino acid sequence <SEQ ID 9926> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAC98439 GB:L29324 UmuC MucB homolog [Streptococcus pneumoniae]
 Identities = 303/436 (69%), Positives = 360/436 (82%)

Query: 39 LHTSLCVMSRADNSAGLILASSPMFKKVFVGKGNVGRAYDLPPDVHTRKFNYYRAKISGLP 98

30 L LCVMSRADNSAGLILASSPMFKKVFVGK NVGR+YDLPPDV TRKF+YY AK GLP

Sbjct: 5 LRLRLCVMSRADNSAGLILASSPMFKKVFGKSNVGRSYDLPPDVKTRKFSYNAKKQGLP 64

Query: 99 TDAKFVSFIENWAKRTFIVPPRMDLYIQKNLEIQKFQNYADPTDILPYSIDEGFIDLTS 158

35 T +V +IE WAK T IVP L I N+EIQK+FQ++A P DI PYSIDEGFIDLTS

Sbjct: 65 TTIDYVRYIEEWAKSTVIVPREWILTIAVNMEIQKIFQDFAAPDDIYPYSIDEGFIDLTS 124

Query: 159 SLNYFVEDKSLSRKDKLDVVSQAKIQHDIWEKTGVYSTVGMNSNANPLAKLALDNEAKTTA 218

40 SLNYFV DKS+SRKDKLD++SA IQ IW KTG+YSTVGMNSNANPLAKLALDNEAK T

Sbjct: 125 SLNYFVPDKSISRKDKLDIISAAIQKKIWRKTGIYSTVGMNSNANPLAKLALDNEAKKTP 184

Query: 219 TMRANWSYEDVETKVWNIPKMTDFWGIGSRTEKRLNKLGIYSIKELANCDPTILKKEFGV 278

45 TMRANWSYEDVE KVV IPKMTDFWGIG+R EKRL+ LGI+SIKELA +P ++KKE G+

Sbjct: 185 TMRANWSYEDVEKKVWTIPKMTDFWGIGNRMEKRLHNLGIFSIKELAQANPDLIKELGI 244

Query: 279 IGVQHWFWHANGIDESNVHEPYRPAKAVGIGNSQVLHKDYTRQSDIELVLEMAEQVAIRLR 338

50 +G++ WFWHANGIDESNVH+PY+PK+ GIGNSQVL KDY +Q DIE++LREMAEQVA+RLR

Sbjct: 245 MGLELWFWHANGIDESNVHVKPKPSKGIGNSQVLPKDYIKQDIEIILREMAEQVAVRLR 304

Query: 339 RRHKKATVVAINVGYSNFENKKSINVQRKINPNRNLTVFQDEVVSLFRSKYDGGAVRSIA 398

55 R KKATVV+I++GYS E K+SIN Q KI P N+T + + V+ LF +KY GA+R++A

Sbjct: 305 RSGKKATVVSIIHLGYSKVEQKRSINTQMKEPTNQTALLTNYVLKLFHTKYTSGAIRNVA 364

Query: 399 VRYDGLVDENFAVISLFDFFESEKEEKELETIDSIRDRFGFLAVQKASSILENSRAISR 458

V Y GLVDE+F +ISLFD E+ EKEE+L++ ID+IR FGF ++ K ++L + SR I+R

Sbjct: 365 VNYSGLVDESFGILISLFDIEKIEKEERLQSAIDAIRTEFGFTSLKGNALDQASRTIAR 424

Query: 459 SRLVGGHSAGGLEGLK 474

S+L+GGHSAGGL+GLK

-2380-

Sbjct: 425 SKLI'GGHSAGGLDGLK 440

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2109

A DNA sequence (GBSx2224) was identified in *S.agalactiae* <SEQ ID 6521> which encodes the amino acid sequence <SEQ ID 6522>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

10

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4016(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2110

A DNA sequence (GBSx2225) was identified in *S.agalactiae* <SEQ ID 6523> which encodes the amino acid sequence <SEQ ID 6524>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

25

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2088(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG13001 GB:AF227520 unknown [Streptococcus pneumoniae]

Identities = 68/122 (55%), Positives = 89/122 (72%), Gaps = 6/122 (4%)

35

Query: 1 MIDRSYLPFKVAREYQDRKMAKWMGFFLSEHTAGLDSELNKVDYTSSELSISDKLLLLNQL 60

MIDRSYLPF+ AREYQD KM KWMGFFLSEHT+ L + NKV Y S+LS+ KLLLL+Q+

Sbjct: 1 MIDRSYLPFQSAREYQDTKMOKWMGFFLSEHTSALTDDANKVTYMSDLSLEKKLLLLSQV 60

40

Query: 61 YSNQLNGIIVPGQ----YYSGKVDNLTFFNHVSLKTKTGFVSIPIKDILSIDL--EVEYE 114

Y+ QLN I V + Y+G + +LT + + +KT TG +++ +KDI+SI+L EV YE

Sbjct: 61 YAGQLNTRIHVVKNNQVSYTGITPSLTGDFILIKTTTGHINLKLKDIVSIELVEEVLVE 120

Query: 115 SA 116

SA

45

Sbjct: 121 SA 122

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2381-

Example 2111

A DNA sequence (GBSx2226) was identified in *S.agalactiae* <SEQ ID 6525> which encodes the amino acid sequence <SEQ ID 6526>. Analysis of this protein sequence reveals the following:

5 Possible site: 48
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4025(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 10 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9927> which encodes amino acid sequence <SEQ ID 9928> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2112

20 A DNA sequence (GBSx2227) was identified in *S.agalactiae* <SEQ ID 6527> which encodes the amino acid sequence <SEQ ID 6528>. This protein is predicted to be soluble transducer HtrXIII. Analysis of this protein sequence reveals the following:

 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5246(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2113

35 A DNA sequence (GBSx2228) was identified in *S.agalactiae* <SEQ ID 6529> which encodes the amino acid sequence <SEQ ID 6530>. Analysis of this protein sequence reveals the following:

 Possible site: 60
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5131(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2382-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2114

A DNA sequence (GBSx2229) was identified in *S.agalactiae* <SEQ ID 6531> which encodes the amino acid sequence <SEQ ID 6532>. This protein is predicted to be pXO2-78. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2105 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF13682 GB:AF188935 pXO2-78 [Bacillus anthracis]
 Identities = 101/314 (32%), Positives = 147/314 (46%), Gaps = 46/314 (14%)

Query: 27 SGQIYEHPDHSFRIFADTNTFKWFSRDIQGDVIDFVQLVAGVSFKKALSYLETG--GFE 84
 S + Y +HDS I N F W SR + G++I FVQ V SF A+ L G +E
 Sbjct: 39 SERYRLTEHDSLIIDRKKNQFYWNSRGVNGNIKFVQEVEDASFPQAMQRLLDGEQDYE 98

Query: 85 EAKVIEETYQPFQYYLREEP----FQQARTYLKDIRGLSNQTINSFGRQGLLAQATYQAE 140
 +A I +P+ Y E+ F +AR YL + R + Q +++ +GL+ Q Y
 Sbjct: 99 KASEITFVSEPYDYEHEFEQKEVSRFDRAREYLIEERKIDPQVVDALHNKGLIKQDKYN-- 156

Query: 141 SVLVFKSFDHNGTLQAASLQGLVKNEEKYDRGYLKKIMKGSFGHVGISFDIGNPKRLIFC 200
 +VL G + S QG+VK++ KY RG K I K S + G + G P+ L F
 Sbjct: 157 NVLFLWKDRETGAVMGGSEQGVVKS-D-KYRGAWKSIQKNSTANYGFNVLNGEPRNLKFY 215

Query: 201 ESVIDMMSYYQLHQQLSDVRLISMEGLKLSVIAYQTIRLAAEEQGKIAFLDTVKPIRLS 260
 ES ID++SY LH+ L D LISMEGLK VI +
 Sbjct: 216 ESDIDLLSYATLHKHNLKDTHLISMEGLKPQVI-----FN 250

Query: 261 HYLQAIQETTTTFQTHSNVITMAVDNDEAGREFYQKL-----SDKGFPIFQ-DLPPLQ 312
 +Y++A + + +++ VDND+AG+ F ++L +D F+ + P
 Sbjct: 251 YYMKACERIGDV---PDSLSLCVDNDKAGKAFVERLIHFRYEKNDGSIVAFKPEYPQAP 306

Query: 313 RLETKSDWNDIVKR 326
 E K DWND KR
 Sbjct: 307 SEEKKWDWNEDECKR 320

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2115

A DNA sequence (GBSx2230) was identified in *S.agalactiae* <SEQ ID 6533> which encodes the amino acid sequence <SEQ ID 6534>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.7013 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2116

A DNA sequence (GBSx2231) was identified in *S.agalactiae* <SEQ ID 6535> which encodes the amino acid sequence <SEQ ID 6536>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1310(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2117

A DNA sequence (GBSx2232) was identified in *S.agalactiae* <SEQ ID 6537> which encodes the amino acid sequence <SEQ ID 6538>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.6726(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9373> which encodes amino acid sequence <SEQ ID 9374> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2118

A DNA sequence (GBSx2233) was identified in *S.agalactiae* <SEQ ID 6539> which encodes the amino acid sequence <SEQ ID 6540>. This protein is predicted to be phosphoglucomutase (manB). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
```

-2384-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2147(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9355> which encodes amino acid sequence <SEQ ID 9356> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB96418 GB:AJ243290 phosphoglucosyltransferase [Streptococcus thermophilus]
 Identities = 391/465 (84%), Positives = 424/465 (91%), Gaps = 1/465 (0%)

Query: 1 MAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA 60
 +A HGIKSYVFE+LRPTPELSFAVRHL+ +AGIM+TASHNPAPFNGYKVYQDGGQ+PPA

15 Sbjct: 107 LAHGIKSYVFESLRPTPELSFAVRHLHTFAGIMITASHNPAPFNGYKVYQDGGQMPPA 166

Query: 61 DADALTDFFIRAIENPFVAVELADLDESKSSGLIQVIGEDVDIEYLREVVDVNNINQDLINNF 120
 DADALTD+IRAI+NPV V+LADL++SK+SGLI++IGE+VD EYL+EVKDVNNINQDLIN +

20 Sbjct: 167 DADALTDYIRAIENPFVAVELADLDESKSSGLIIEIIGENVDAEYLKEVVDVNNINQDLINNEY 226

Query: 121 GKDMKIVYTPHGTGEMLTRRALAQAGFESVSVVSVQAKADPDFSTVKSPNPESQAALFAL 180
 G+DMKIVYT LHGTGEML RRALAQAGF++V VVE+QA DF TVKSPNPES+Q AFAL

25 Sbjct: 227 GRDMKIVYTPHGTGEMLVRRALAQAGFDAVQVVEAQAVPHADFLTVKSPNPENQDAFAL 286

Query: 181 AEEELGREVDADVIVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL 240
 AEEELGR VDADVIVATDPDADRLGVEIRQPDGSY NLSGNQIGAIIAKYILEAHKTAGTL

30 Sbjct: 287 AEEELGRNVDAVIVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL 346

Query: 241 PENAALAKSIVSTELVTIKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGEESF 300
 P NAAL KSIVSTELVTIKIAESYGATMFNVLTGFKFI EKI EFE +HN+TYMFGEESF

35 Sbjct: 347 PANAALCKSIVSTELVTIKIAESYGATMFNVLTGFKFIEKIHEFETQHNYTYMFGEESF 406

Query: 301 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 360
 GYLIKPFVRDKDAIQAVL+VAEIAAYYRSRG+TLADGI+EIYK+YGYF+EKTISVTLSGV

40 Sbjct: 407 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIEIYKQYGYFSEKTISVTLSGV 466

Query: 361 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTLA 420
 DGAAEIKKIMDKFR N PKQFNNTDI EDF +QTAT DG + LITPPSNVLYK LA

45 Sbjct: 467 DGAAEIKKIMDKFRNAPKQFNNTDIKTEDFLEQTATTADG-VEKLTTPPSNVLYKILA 525

Query: 421 DDSWIAVRPSGTEPKIKFYIATVGNLADAETKIANIEKEITTFV 465
 DDSW AVRPSGTEPKIKFYIATVG ADA+ KIANIE EI FV

50 Sbjct: 526 DDSWFAVRPSGTEPKIKFYIATVGETEADAKEKIANIEAEINAFV 570

45 There is also homology to SEQ ID 6156:

Query: 1 MAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA 60
 +AQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA

50 Sbjct: 107 LAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA 166

Query: 61 DADALTDFFIRAIENPFVAVELADLDESKSSGLIQVIGEDVDIEYLREVVDVNNINQDLINNF 120
 DADALTDFFIRAIENPFVAVELADLDE+KSSGLIQVIGEDVD+EYLREVVDVNNINQDLINNF

55 Sbjct: 167 DADALTDFFIRAIENPFVAVELADLDEKSSGLIQVIGEDVDMEYLREVVDVNNINQDLINNF 226

Query: 121 GKDMKIVYTPHGTGEMLTRRALAQAGFESVSVVSVQAKADPDFSTVKSPNPESQAALFAL 180
 GKDMKIVYTPHGTGEMLTRRALAQAGFESVSVVSVQAKADPDFSTVKSPNPESQAALFAL

60 Sbjct: 227 GKDMKIVYTPHGTGEMLVRRALAQAGFESVSVVSVQAKADPDFSTVKSPNPESQAALFAL 286

Query: 181 AEEELGREVDADVIVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL 240
 AEEELGREV+ADVIVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL

65 Sbjct: 287 AEEELGREVADVIVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL 346

Query: 241 PENAALAKSIVSTELVTIKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGEESF 300
 PENAALAKSIVSTELVTIKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGEESF

70 Sbjct: 347 PENAALAKSIVSTELVTIKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGEESF 406

-2385-

Query: 301 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 360
 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV
 5 Sbjet: 407 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 466

Query: 361 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA 420
 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA
 Sbjet: 467 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA 526

10 Query: 421 DDSWIAVRPSGTEPKIKFYIATVGNDLADAETKIANIEKEITTFV 465
 DDSWIAVRPSGTEPKIKFYIAT+G+ L A+ KIANIE EI TFV
 Sbjet: 527 DDSWIAVRPSGTEPKIKFYIATIGDTLDIAQEKIANIETEINTFV 571

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 2119

A DNA sequence (GBSx2235) was identified in *S.agalactiae* <SEQ ID 6541> which encodes the amino acid sequence <SEQ ID 6542>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1564 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9905> which encodes amino acid sequence <SEQ ID 9906> was also identified. There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 2120

A DNA sequence (GBSx2236) was identified in *S.agalactiae* <SEQ ID 6543> which encodes the amino acid sequence <SEQ ID 6544>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.92	Transmembrane	162 - 178 (135 - 184)
INTEGRAL	Likelihood = -7.11	Transmembrane	58 - 74 (56 - 78)
INTEGRAL	Likelihood = -6.42	Transmembrane	136 - 152 (135 - 161)
40 INTEGRAL	Likelihood = -5.20	Transmembrane	23 - 39 (21 - 49)
INTEGRAL	Likelihood = -1.75	Transmembrane	485 - 501 (485 - 501)

----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35376 GB:AB001710 ABC transporter, ATP-binding protein
 50 [Thermotoga maritima]
 Identities = 216/552 (39%), Positives = 336/552 (60%), Gaps = 3/552 (0%)

Query: 26 MALLGTVVQVCLTVYLPVLIGQAVDVVLSPHSMILLPLIMWKMIIVILANTTIQWINPLL 85
 M + V L V P LIG+ +DVV P LL M + + +++ W+ +

-2386-

Sbjct: 41 MVFVFTVSSILGVLSPYLIGKTIDVVFVPRRFDLLPRYMLILGTIYALTSLFLWLQGGKI 100

Query: 86 YNRLIFHYVASLRKAVMEKLNLLPIAYLDKRGIGDLISRVTDTQELSNGLLMVFNQFFV 145
 L V LRK + EKL +P+ + D+ GD+ISRV D + ++N L QFF

5 Sbjct: 101 MLTSLQDVVFRRLRKELFEKLQRPVPGFFDRTPHGDIISRVDNDNINNVLGNSIIQFFS 160

Query: 146 GLLTIIVTIFSMKIDLLMLFLVLFLTPLSLFLARFIKKSYS-HLYQNQTASRGRQTQFI 204
 G++T+ + M +++++ + L + PL++ + + ++ + + Y+NQ G+ I

10 Sbjct: 161 GIVTLAGAVIMMFRVNVILSLVTLIVPLTVLITQIVSSQTRKYFYENQVRVL-GQLNGII 219

Query: 205 EEMVSQESLIQAFAQESSDHFRTINQEYANFSQSAIFYSSVTNPSTRFINSIYGFILA 264
 EE +S ++I+ F+ +E+ + F +N+ A +S + P +N+L + ++

Sbjct: 220 EEDISGLTVIKLFTREEKEMEKFDVNESLRKVGTKAQIFSGVLPPLMMVNNLGFALIS 279

15 Query: 265 GIGALRIMSGAFSVGQLITFLNYVNQYTKPFNDISSVLSEMQSALACAERLYSILEESSP 324
 G G + +VG + TF+ Y Q+T+P N++S+ + +Q ALA AER++ IL+

Sbjct: 280 GFGGWLAKDITVGTIATFIGYSRQFTPLNELSNQFNMIQMALASAEIRIFEILDLEEE 339

Query: 325 NITGTEKLDSSSTVKGQIDFKNVVFGYNKSKLLNGLNHLHIPAGAKVAIVGPTGAGKSTLI 384
 + ++ V+G+I+FKNV F Y+K K +L I HI G KVA+VGPTG+GK+T++

20 Sbjct: 340 K-DDPDAREVLRGEIEFKNVWFSYDKKKPVLKIDITFHIKPGQKVALVGPTGSGKTTIV 398

Query: 385 NLIMRFYEVDGGNILLDCKPITDYEPSQLRQEI GMVLQETWLKSATIHNDNIAYANPKASR 444
 NL+MRFY+VD G IL+D I + S LR IG+VLQ+T L S T+ +N+ Y NP A+

25 Sbjct: 399 NLLMRFYDVRGQILVDGIDIRKIKRSSLRSSIGIVLQDTILFSTTVKENLKYGNPGATD 458

Query: 445 EEVIEAAKAANADFFIKQLPNGYDYLEADAGDSLSQGCQLLTARIIFLKLPRILILDEA 504
 EE+ EAAK ++D FIK LP GY+T L D G+ LSQGG QLL I R FL P+ILILDEA

30 Sbjct: 459 BEIKEAAKLTSDHFIKHLPEGYETVLTNDGEDLSQGRQLLAITRAFLANPKILILDEA 518

Query: 505 TSSIDTRTEVLVQEAQMLMKGRSTFIIAHLSTIQTADIILVMVSGEIVEVGNHSELMA 564
 TS++DT+TE +Q A LM+G+TS IIAHRL+TI+ AD+I+V+ GEIVE+G H EL+

Sbjct: 519 TSNVDTKTEKSIQAAMWKLMEGKTSIIIAHRLNTIKNADLIIVLRDGEIVEMGKHDELIQ 578

35 Query: 565 QKGIYYQMNAQ 576
 ++G YY++ +Q

Sbjct: 579 KRGFYELFTSQ 590

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6545> which encodes the amino acid
 40 sequence <SEQ ID 6546>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.07	Transmembrane	162 - 178 (159 - 182)
INTEGRAL	Likelihood = -7.17	Transmembrane	143 - 159 (137 - 161)
45 INTEGRAL	Likelihood = -5.84	Transmembrane	23 - 39 (19 - 45)
INTEGRAL	Likelihood = -5.68	Transmembrane	68 - 84 (60 - 86)
INTEGRAL	Likelihood = -2.55	Transmembrane	261 - 277 (256 - 278)

----- Final Results -----

50 bacterial membrane --- Certainty=0.4227(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 >GP:AAD35376 GB:AE001710 ABC transporter, ATP-binding protein
 [Thermotoga maritima]
 Identities = 206/572 (36%), Positives = 342/572 (59%), Gaps = 5/572 (0%)

Query: 2 IKTDHLLKRVLDLLKKPLPVCILVIAFVQVG--LSVYLPVLIGKAVDMSLSVNSWQT 59
 +K L+R+L L +P +++++ FV V L V P LIGK +D+ +

60 Sbjct: 18 LKNPTATLRLRLGYL--RPHTFTLIMVVFVFTVSSILGVLSPYLIGKTIDVVFVPRRFDL 75

Query: 60 LKWLGLQMLVIVVNTLIQWVMPVLSRLLYQYSQQLKDKLLEKIHRLPFAYLDRQTIGD 119
 L + + I + +L+ W+ + L +L+ +L EK+ R+P + DR GD

65 Sbjct: 76 LPRYMLILGTIYALTSLFLWLQGGKIMLTSLQDVVFRRLRKELFEKLQRPVPGFFDRTPHGD 135

-2387-

5 Query: 120 LVSRVITDTEQLINGLQMFVNQFILGLLTILCTIIAMAQIDWMLLILVLVLTSSFLAR 179
 ++SRVI D + + N L QF G++T+ +I M +++ ++ ++ L + P ++ + +
 Sbjct: 136 IISRVINDVDNINNVLGNSIIQFFSGIVTLGAVIMMFRVNVILSLVTLISIVPLTVLITQ 195

10 Query: 180 FIAQKSFHYAQAQTKSRGNLAQFTEBILRQEGVLQLFNAQEQSICDYHVLNKTYCEASQK 239
 ++ ++ Y + G L EE + +++LF +E+ + + +N++ + K
 Sbjct: 196 IVSSQTRKYFYENQRVLGQLNGIIBEDISGLTVIKLFTREBEMEKFDORVNESLRKVGTK 255

15 Query: 240 AIFYASTVNPATRFINSVIYALLAGLGAVRIMAGLFSVGQLTTFNLVVVQYTKPFNDISS 299
 A ++ + P +N++ +AL++G G + + +VG + TF+ Q+T+P N++S+
 Sbjct: 256 AQIFSGVLPLPLMMVNNLGFALISGFGWLALKDITVTGTIATFIGYSRQFTRPLNELSN 315

20 Query: 300 VLAEIQSSLACQRLYDLDDIEIKEQEHLTFKASAVKGQIDFEEVSFSYQKDRPLLKDI 359
 IQ +LA A+R++++LD+E +E++ + V+G+I+F+ V FSY K +P+LKDI
 Sbjct: 316 QFNMIQMALASABRIFEILDLE-EEKDDPDAVELREVRGBIEFKNVWFSYDKKKPVLKDI 374

25 Query: 360 NFSVPAGSKVAIVGPTGAGKSTLINLLMRFYELDAGSIKLDKVPKCYAKEELRSITGIV 419
 F + G KVA+VGPTG+GK+T++NLLMRFY++D G I +D + I+ + LRS GIV
 Sbjct: 375 TFHIKPGQKVALVGPTGSGKTTIVNLLMRFYDVRGQILVDGIDIRKIKRSSLSRSIGIV 434

30 Query: 420 LQETWLKDATVHELIAYGSEASRDEVVAAAKAAHAHFFIMQLPKTYDTYLSASDDALSQ 479
 LQ+T L TV E + YG+ A+ +E+ AAK H+ FI LP+ Y+T L+ + + LSQ
 Sbjct: 435 LQDTILFSTTVKENLKYGNPGATDEEIKEAAKLTHSDHFIKHLPEGYETVLTDNEDLSQ 494

35 Query: 480 QQLQLLAIARMFLKKPKVLVLDEATSSIDIRTEAVIQEALKELMRGRTSFIIAHLSTIQ 539
 GQ QLLAI R FL PK+L+LDEATS++D +TE IQ A+ +LM G+TS IIAHL+TI+
 Sbjct: 495 GQRQLLAI TRAFLANPKILILDEATS NVDTKTEKSIQAAMWKLMEGKTSIIIAHLNLTIK 554

40 Query: 540 SADLILVMDQGRIVEWGTTHASIMSKNGCYVRL 571
 +ADLI+V+ G +VE G H L+ K G Y L
 Sbjct: 555 NADLIIVLRDGEIVEMGKHDELIQKRGFYEL 586

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 340/566 (60%), Positives = 433/566 (76%)

40 Query: 11 KKLVDLLSKKSLVGMALLGTVVQVCLTVYLPVLIGQAVDVVLSPHSMILLPLIMWKMIA 70
 K+++QDLL K V + ++ + VQV L+VYLPVLIG+AVD+ LS +S L ++ +M+
 Sbjct: 10 KRVLQDLLKKPLPVCLLVIA SFVQVGLSVYLPVLIGKAVDMSLSVNSWQTLKWLGLGQMLV 69

45 Query: 71 VILANTIIQWINPLLYNRLIFHYVASLRKAVMEKLNLLPIAYLDKRGIGDLISRVTTDTE 130
 +I+ NT+IQW+ PL+Y+RL++ Y L+ ++EK++ LP AYLD++ IGDL+SRV TDTE
 Sbjct: 70 IIVVNTLIQWVMPVYSRLLYQYSQQLKDKLLEKIHRLPFAYLDRQTIGDLVSRVITDTE 129

50 Query: 131 QLSNGLLMVFVNQFFVGLLTIIIVTIFSMKIDLLMLFLVLFLTPLSLFLARFIAKSYHLY 190
 QL NGL MVFNQF +GLLTI+ TI +MA+ID LML LVL LTP SLFLARFIA+KS+H
 Sbjct: 130 QLINGLQMFVNQFILGLLTILCTIIAMAQIDWMLLILVLVLTSSFLARFIAQKSFHYA 189

55 Query: 191 CNQTASRGRQTQFIEEMVSQESLIQAFAQEESDHFRTINQEYANFSQSAIFYSTVNP 250
 Q QT SRG QF EE++ QE L+Q F+AQE+S + +N+ Y SQ AIFY+STVNP
 Sbjct: 190 QAQTKSRGNLAQFTEBILRQEGVLQLFNAQEQSICDYHVLNKTYCEASQKAIIFYASTVNP 249

60 Query: 251 STRFINSIIYGFAGIGALRIMSGAFSVGQLITFLNVVNQYTKPFNDISSVLSEMQSALA 310
 +TRFINS+IY LAG+GA+RIM+G FSVGQL TFLN V QYTKPFNDISSVL+E+QS+LA
 Sbjct: 250 ATRFINSVIYALLAGLGAVRIMAGLFSVGQLTTFNLVVVQYTKPFNDISSVLAEIQSSLA 309

65 Query: 311 CAERLYSILEESSPNITGTEKLDSSSTVKQIDFKNVVFGYNSKILLNGINLHIPAGAKV 370
 CA+RLY +L+ +S VKGQIDF+ V F Y K + LL IN +PAG+KV
 Sbjct: 310 CAQRLYDLDDIEIKEQEHLTFKASAVKGQIDFEEVSFSYQKDRPLLDINFSVPAGSKV 369

70 Query: 371 AIVGPTGAGKSTLINLIMRFYEVDDGNILLDCPKPITDYEPSQLRQEIGMVLQETWLKSAT 430
 AIVGPTGAGKSTLINL+MRFYE+D G+I LD PI Y +LR G+VLQETWLK AT
 Sbjct: 370 AIVGPTGAGKSTLINLLMRFYELDAGSIKLDKVPKCYAKEELRSITGIVLQETWLKDAT 429

75 Query: 431 IHDNIAYANPKASREEVIEAAKAANADFFIKQLPNGYDTYLEDAGDSLSQGCQLLTIAR 490
 +H+ IAY + +ASR+EV+ AAKAA+A FFI QLP YDTYL + D+LSQGC QLL IAR
 Sbjct: 430 VHELIAYGSEASRDEVVAAAKAAHAHFFIMQLPKTYDTYLSASDDALSQGCQLQLLAIAR 489

-2388-

Query: 491 IFLKLPRIILDEATSSIDTRTEVLVQEAQFQMLMGRTSFIIAHLSTIQTADILVMVS 550
 +FLK P++L+LDEATSSID RTE ++QEA + LM+GRTSFIIAHLSTIQ+AD+ILVM
 Sbjct: 490 MFLKKPKVLVLDEATSSIDIRTEAVIQEALKELMRGRTSFIIAHLSTIQSADLILVMDQ 549

Query: 551 GEIVEVGNHSELMAQKGIYYQMNAQ 575
 G +VE G H+ LM++ G Y ++Q +
 Sbjct: 550 GRLVEWGTASLMSKNGCYVRLQKIE 575

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2121

A DNA sequence (GBSx2237) was identified in *S.agalactiae* <SEQ ID 6547> which encodes the amino acid sequence <SEQ ID 6548>. Analysis of this protein sequence reveals the following:

15 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1099(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2122

A DNA sequence (GBSx2238) was identified in *S.agalactiae* <SEQ ID 6549> which encodes the amino acid sequence <SEQ ID 6550>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

30 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -13.69	Transmembrane	157 - 173 (130 - 182)
INTEGRAL	Likelihood = -10.88	Transmembrane	56 - 72 (49 - 77)
INTEGRAL	Likelihood = -7.75	Transmembrane	239 - 255 (235 - 258)
INTEGRAL	Likelihood = -6.42	Transmembrane	133 - 149 (130 - 156)
INTEGRAL	Likelihood = -4.78	Transmembrane	271 - 287 (270 - 289)
INTEGRAL	Likelihood = -1.91	Transmembrane	20 - 36 (20 - 37)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.6477(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35375 GB:AE001710 ABC transporter, ATP-binding protein
 [Thermotoga maritima]
 Identities = 196/570 (34%), Positives = 327/570 (56%), Gaps = 5/570 (0%)

50 Query: 1 MKRLTYFFKGYIKETIFGPLFKLLEASFELLVPPIVIAKMIDETIPRGDRSGLLLQIGLIF 60
 MK L Y K Y + PLF ++E +L P ++A+++DE I RGD S L+L+ G++
 Sbjct: 1 MKTLARYLKPYWIFAVLAPLFMVVEVICDLSOPTLLARIVDEGIARGDFS-LVLKTGILM 59

Query: 61 FLAA-VGVVVAITAQYYSSKAAVGYTRQLTEDLYQKVMISLGKIDRDELGTASLITRLTAD 119

-2389-

```

      + A +G V I  ++S A+  +  L DL++KV+S  + +  T+SLITRLT D
Sbjct: 60  LIVALIGAVGGIGCTVFASYASQNFADLRDLFRKVLSPSISNVNRFHTSSLITRLTND 119

5  Query: 120  TFQIQTLGNQFLRLFLRAPIIIVFGAIIMAFSISPSLTIWFLVMVVTLLFIIVFVMSRLINP 179
      Q+Q  +  LR+ +RAP++  G I+MA SI+  L+  + ++  + ++  +++  NP
Sbjct: 120  VTQLQNLVMMLLRIVVRAPLLFVGGIVMAVSINVKLSSVLIFLIPPVLLFVWLTKKGNP 179

10 Query: 180  IYLKIRTSTDYLVKLTRQQLQGVVRIRAFNQVDRESEAFNDINYHTNLQKAGRLSSLV 239
      ++ KI+ STD + ++ R+ L GVRV+RAF + + E+E F  N  + A L
Sbjct: 180  LFRKIQESTDEVNRVRENLLGVRVVRAFRREEYENENFRKANESLRRSIISAFSLIVFA 239

15 Query: 240  TPLTFLVVNITLVVLIWRGNLNIANHLLSQGMLVALINYLQILVELLKMIMLVTSINQS 299
      PL  +VN+ ++ ++W G + + N+ +  G ++A NYL+QI+  L+ + ++  + ++
Sbjct: 240  LPLFIFIVNMGMIAVLWFGVLRNQMIEIGSIMAYTNYLMQIMFSLMIGNILNFIVRA 299

20 Query: 300  YISAKRIIAVF-ERPS-EIIDDKLEPKYSNKALEVQEMAFSYPNSSEKALSDITFSMNVG 357
      SAKR++ V  E+P+ E  D+ L  ++  + + F Y  +++  LS + FS+  G
Sbjct: 300  SASAKRVLEVLNEKPAIEADNALALPNVEGSVSFENVEFRYFENTDPVLSGVNFSVKPG 359

25 Query: 358  ETLGIIGGTGSGKSTLINLLHHYKVQEGDIDIYHQKSPDTISNWRTLVRVVPQNAQLF 417
      + ++G TSGKSTL+NL+  +  + G +++  + + R  +  VPQ  LF
Sbjct: 360  SLVAVLGETGSGKSTLMNLIPRLIDPERGRVEVDLDRVTVKLKDRLRGHISAVPQETVLF 419

30 Query: 418  KGTIRSNLSLGLGKVSEELWTALEIAQASDFVKEKDQGLDAPVESFGRNFGGQQRRLT 477
      GTI+ NL  G  +++++  A +IAQ  DF+  D+ VE  GRNFGGQ+QRL+
Sbjct: 420  SGTIKENLKWGREDATDDEIVEAAKIAQIHDFIISLPEGYDSRVERGGRNFGGQQRRLS 479

35 Query: 478  IARALVQDKIPFLILDDATSALDYLTEARLFKAITKHFNQTNLIIVSQRINSIQNADRIL 537
      IARALV+ K  LILDD TS++D +TE R+  + ++  I++Q+I +  AD+IL
Sbjct: 480  IARALVK-KPKVLILDDCTSSVDPITEKRILDGLKRYTKGCTTFIITQKIPTALLADKIL 538

Query: 538  LLDKKGKQVGFDNHQSLLAHNKVYKSIYHSQ 567
      +L +GK  GF  H+ LL  H  K  Y+  IY  SQ
Sbjct: 539  VLHEGKVAGFGTHKELLEHCCKPYREIYESQ 568

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6551> which encodes the amino acid sequence <SEQ ID 6552>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence
40  INTEGRAL  Likelihood = -12.47  Transmembrane  157 - 173 ( 149 - 185)
    INTEGRAL  Likelihood = -7.75  Transmembrane  55 - 71 ( 51 - 74)
    INTEGRAL  Likelihood = -4.25  Transmembrane  239 - 255 ( 237 - 260)
    INTEGRAL  Likelihood = -3.77  Transmembrane  20 - 36 ( 19 - 37)
    INTEGRAL  Likelihood = -3.50  Transmembrane  271 - 287 ( 270 - 288)
45  INTEGRAL  Likelihood = -2.55  Transmembrane  133 - 149 ( 130 - 151)

----- Final Results -----
      bacterial membrane --- Certainty=0.5989(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

1GB:AL137187 putative ABC transporter [Streptomyces ... 296 6e-79
55 >GP:CAB69751 GB:AL137187 putative ABC transporter [Streptomyces
    coelicolor A3(2)]
    Identities = 185/569 (32%), Positives = 306/569 (53%), Gaps = 8/569 (1%)

60 Query: 1  MKRLRPYVKGYLKESILGPLFKLLEALFELLVPLLIANMIDISISQHNSQGILRVVLTFL 60
      ++ LR Y++ Y K  L  +L+  L +P L A++ID + + +S  IL  +
Sbjct: 3  IRLRRTYLRPYKKPIALLVALQFLQTCASLYLPTLNAHIIDEGVVKGDSGYILSYGALMI 62

Query: 61  GLATIGLLLSVTAQYFSSKAAVGFTRQMTDDLFPKIMFLSKEDQDHLGYASLLSRLTSDS 120
      G++  ++ ++ A ++ ++ A  R +  +F ++  S  +  H G  SL++R T+D
65 Sbjct: 63  GISLAQVVCNIGAVFYGARTAAALGRDVRGAVFDRVQSFSAREVGHFGAPSLITRTTNDV 122

```

-2390-

5 Query: 121 FQIQTGINQFLRLFLRAPII VCGAMVMAYWISPSLTLWFVMMVIVLLTLVFMVSHLLGPL 180
 Q+Q L + API+ G +VMA + L+ + +V VL V ++ L PL
 Sbjct: 123 QQVQMLALMTFTILMVSAPI MCVGGIVMALGLDVPLSGVLLGVVPVLAICVTLIVRKLRL 182

10 Query: 181 YLLIRRETDLHVLRLTSQQLOGIRVIKAFNQTKELQAFKQONMLLSRHQYQAATLANVLN 240
 + ++ D + R+ +Q+ G RVI+AF + + E Q F++ N L+ L ++
 Sbjct: 183 FRKMQVRLDTVNRVLRQITGNRVIRAFVRDEYEQQFRKANTELTEVALGTGNLLALMF 242

15 Query: 241 PMTFLVVNLTLILLIWQGSQVAHRSLSQGMVALINYLQILAEELLKMTMLMGTINQSV 300
 P+ VVNL+ + ++W G+ ++ + G L A + YL+QI+ ++ T + + ++
 Sbjct: 243 PVVMTVVNLSIAVVWFGAHRIDSGGMQIGDLTAFLAYLMQIVMSVMMATFMFMMPRAE 302

20 Query: 301 TAAKRINQVFVLADEAPLPLKDGPISTH-LLTIRHLTFTYPGAAEPSLYDIQLSADQGE 359
 A+RI +V P+ + H L IR F YPGA EP L I L A GE
 Sbjct: 303 VCAERIQEVLETSSVVPVAPVTELRRHGHLEIREAGFRYPGAEPPVLRHIDLVARPGE 362

25 Query: 360 WIGIIGGTGAGKTTLIDLICQTSQYSQSGEISLNW---QGEVPKLTTEWRNVIALVLPQKAQ 416
 +IG TG+GK+TL+ L+ + + GE+ +N + PKTL + V++LVPQK
 Sbjct: 363 TTAIVIGSTGSGKSTLLGLVPRLFDATDGEVLVNGVDVRTVDPKTLAK---VVSILVPQKPY 419

30 Query: 417 LFKGTIRSNLLLGQSMPISEELWRALELAQAKEFVAALPEQLEAPVEAFGRHFSGGQRQ 476
 LF GT+ +NL G + +DEELW AL +AQAKEFV+ L L+AP+ G + SGGQRQ
 Sbjct: 420 LFAGTVATNLRYG-NPDATDEELWHA LAQAKEFVSELEGGDLAPIAQGGTNVSGGQRQ 478

35 Query: 477 RLAIARALLKPKPILILDDASSALDNETRGRLEFKALKEELSDVLVILVTQSIKNLQFADK 536
 RLAIAR L++ I + DD+ SALD T L L +E ++ V++V Q + ++ AD+
 Sbjct: 479 RLAIARTLVQRPEIYLFDDSFSA LDYATDAALRAELAQETAETVIVVIAQRVATIRADR 538

40 Query: 537 ILVLEQGHQLDFASHDQLKVSNALYQEML 565
 I+VL++G + H +L N Y+E++
 Sbjct: 539 IIVLDEGRVVGVRHHELMADNETYREIV 567

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 313/568 (55%), Positives = 428/568 (75%), Gaps = 9/568 (1%)

40 Query: 1 MKRLTYYPKGYIKETIFGPLFKLLEASFELLVPIVIAKMIDETIPRGDRSGLLLQIGLIF 60
 MKRL Y KGY+KE+I GPLFKLLEA FELLVP++IA MID +I + + G+L + +F
 Sbjct: 1 MKRLRPYVKGYLKEISILGPLFKLLEALFELLVPLLIANMIDISISQHN SQILRVVLTIF 60

45 Query: 61 FLAAGVVVAITAQYYSKAAVGYTRQLTEDLYQKVMISLGKKDRDELGTASLITRLTADT 120
 LA +G+++++TAQY+SSKAAVG+TRQ+T+DL++K+M L K+D+D IG ASL++RLT+D+
 Sbjct: 61 GLATIGLLLSVTAQYFSSKAAVGFTQMTDDLFPKIMFLSKEDQDHLGYASLLSRLTSDS 120

50 Query: 121 FQIQTGINQFLRLFLRAPII VFGAIIAFSISPSLTIWFLVMVVTLFIIVFVMSRLLNPI 180
 FQIQTG+NQFLRLFLRAPII V GA++MA+ ISPSLT+WF++MV+ L +VFMV LL P+
 Sbjct: 121 FQIQTGINQFLRLFLRAPII VCGAMVMAYWISPSLTLWFVMMVIVLLTLVFMVSHLLGFL 180

55 Query: 181 YLKIRSTDYLVKLTRQQLOGVRVIRAFNQVDRESEAFNDINYHYTNLQKAGRLSSSLVT 240
 YL IR TD+LV+LT QQLQG+RVI+AFNQ +E +AF N + Q +A L++++
 Sbjct: 181 YLLIRRETDLHVLRLTSQQLOGIRVIKAFNQTKELQAFKQONMLLSRHQYQAATLANVLN 240

60 Query: 241 PLTFLVVNITLVVVIWRGNLNIANHLLSQGMVALINYLQILVELLKM TMLVTSINQSY 300
 P+TFLVVN+TL+++IW+G+ +A+ LSQGMVALINYLQIL ELLKM TML+ ++NQS
 Sbjct: 241 PMTFLVVNLTLILLIWQGSQVAHRSLSQGMVALINYLQILAEELLKMTMLMGTINQSV 300

65 Query: 301 ISAKRIIAVF---ERPSEIIDDKLEPKYSNKALEVQEMAFSYPNSSEKALS DITFSMNV 356
 +AKRI VF E P ++ D S L ++ + F+YP ++E +L DI S +
 Sbjct: 301 TAAKRINQVFVLADEAPLPLKDGPISTHLLTIRHLTFTYPGAAEPSLYDIQLSADQ 357

Query: 357 GETLGIIGGTGSGKSTLINLLHIYKVQEGDIDIYHQGKSPDTISNWR TLVRVVPQNAQL 416
 GE +GIIGGTG+GK+TLI+L+ Y G+I + QG+ P T++ WR ++ +VPQ AQL
 Sbjct: 358 GEWIGIIGGTGAGKTTLIDLICQTSQYSQSGEISLNWQGEVPKLTTEWRNVIALVLPQKAQL 417

Query: 417 FKGTTIRSNLSLGLG-KVSEEKLWTALEIAQASDFVKEKDGQLDAPVESFGRNFSGGQRQR 475
 FKGTTIRSNL LG +S+E+LW ALE+AQA +FV QL+APVE+FGR+FSGGQRQR
 Sbjct: 418 FKGTTIRSNLLGQSMPISEELWRALELAQAKEFVAALPEQLEAPVEAFGRHFSGGQRQR 477

-2391-

Query: 476 LTIARALVQDKIPFLILDDATSALDYLTEARLFKAITKHFNQTNLIIVSQRINSIQNADR 535
 L IARAL++ K P LILDDA+SALD T RLFKA+ + + +I+V+Q I ++Q AD+
 Sbjet: 478 LAIARALLKPK-PILILDDASSALDNETRGRILFKALKEELSDVLVILVTQSIKNLQFADK 536

Query: 536 ILLLDKKGQVGFNDHQSLLAHNKVYKSI 563
 IL+L++G Q+ F +H L N +Y+ +
 Sbjet: 537 ILVLEQGHQLDFASHDQLKVSNALYQEM 564

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2123

A DNA sequence (GBSx2239) was identified in *S.agalactiae* <SEQ ID 6553> which encodes the amino acid sequence <SEQ ID 6554>. Analysis of this protein sequence reveals the following:

15 Possible site: 43
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.26 Transmembrane 8 - 24 (1 - 28)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAB84433 GB:AF027868 RAS-related protein [Bacillus subtilis]
 Identities = 53/140 (37%), Positives = 78/140 (54%), Gaps = 2/140 (1%)

Query: 28 VKKVLQYHDLVQNTLAENGSEANVHLVLSMIYTETKGAIDVMQSSSISGTTNSITDSH 87
 ++++ Y LV+ L G L+L M+Y E+KG D MQSSES+ N ITD
 30 Sbjet: 49 LERLTDYKPLVEEELSQGLSNYTSILGMMYQESKKGNDPMQSSSISGLKRNETTDPQ 108

Query: 88 TSIKHGVTLTLLSQNISQAKKAKVDVWTAQYAYNFGSSYIDYVADHGGENSEIELAKNYSKNV 147
 S+K G+ + K+ VD+ T +Q+YN G+ YID+VA+HGG ++ ELAK YS+
 35 Sbjet: 109 LSVKQGIKQFTLMYKTKGKGVLDLTIIQSYNMGAGYIDFVAEHGGTHTEELAKQYSEQQ 168

Query: 148 VA--PSLGNNGDTYFYYP 165
 V P L G+ + +P
 Sbjet: 169 VKKNPDLYTCGGNAKNFRYP 188

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4143> which encodes the amino acid sequence <SEQ ID 4144>. Analysis of this protein sequence reveals the following:

45 Possible site: 42
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.66 Transmembrane 8 - 24 (7 - 25)

----- Final Results -----
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 134/200 (67%), Positives = 165/200 (82%), Gaps = 1/200 (0%)

55 Query: 1 MFKFLKRLIALIIIIIFIGYRLVIIHENVKKVLQYHDLVQNTLAENGSEANVHLVLSMIYT 60
 MF+ LKR + +++ F+ Y+ +IH NV++VL Y +V+ TLAEN ++ANV LVL+MIYT
 Sbjet: 1 MFRLLKRACSFLLL-FVIYQSFVIHNNVQRVLAYKPMVEKTLAENDTKANVDLVLAMIYT 59

Query: 61 ETKGDAIDVMQSSSISGTTNSITDSHTSIKHGVTLTLLSQNISQAKKAKVDVWTAQYAYNF 120
 ETKG DVMQSSS SG NSITDS SI+HGV LLS N++ A++A VD WTAVQAYNF

-2392-

Sbjct: 60 ETKGGEADVMOSSSESSSGQKNSITDSQASIEHGVNLLSHNLALAEAGVDSWTAVQAYNF 119

Query: 121 GSSYIDYVADHGGENSELAKNYSKNVAPSLGNYNGDTYFYHPLALISGGKLYKNGCN 180
G++YIDY+A+HGG+N+++LA YSK VVAPSLGN +G TYFYHPLALISGGKLYKNGCN

Sbjct: 120 GTAYIDYIAEHGGQNTVDLATTYSKTVVAPSLGNTSGQTYFYHPLALISGGKLYKNGCN 179

Query: 181 IYYSREVQFNLYLIKIMELF 200

IYYSREV FNLYLI++M LF

Sbjct: 180 IYYSREVFHFNLYLIELMSLF 199

SEQ ID 6554 (GBS244) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 4; MW 23.1kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 2; MW 48kDa).

GBS244-GST was purified as shown in Figure 211, lane 5.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2124

A DNA sequence (GBSx2240) was identified in *S. agalactiae* <SEQ ID 6555> which encodes the amino acid sequence <SEQ ID 6556>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2401(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9837> which encodes amino acid sequence <SEQ ID 9838> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB71302 GB:AJ130879 hypothetical protein [Clostridium
sticklandii]

Identities = 32/95 (33%), Positives = 53/95 (55%), Gaps = 1/95 (1%)

Query: 235 LSPEKLADQLFDDNLTLARLTFVDELKDAIPGPVQVSDIDHSRQIKKENQKLSLSNGIEL 294

LS EK + F++ + + + L A Q+ ++ + +K E QK+ +GIE+

Sbjct: 2 LSVEKALETAFEETDEIKATYKEALSKAGIENEQI-EVSETALKRKFEIQKIITESGIEV 60

Query: 295 IVPNNVYQDAESVEFIQNPDTYSILIKNIQDIQN 329

+P N Y D +EF+ N DGT S++IKNI +IQ+

Sbjct: 61 KIPVNYGDPKLEFVANGDGTVSLVIKNIGNIQS 95

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6557> which encodes the amino acid sequence <SEQ ID 6558>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3336(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2393-

Identities = 246/325 (75%), Positives = 286/325 (87%)

Query: 6 MMDFYIKQIIHQFSPNDTELVLSDTPLTLTPRIDYFRKKLSKVFSDEAKRGYFGEDNV 65
 M+D YIK+I+IHQFSPNDTEL+LSD +++TPRID+YFRKKL+KVFSDEAKRG F +N
 5 Sbjct: 1 MLDSYIKRIVIHQFSPNDTELLSLDRLVSITPRIDEYFRKKLAKVFSDEAKRGQFEANNT 60

Query: 66 FMSHLQDDLYVSSCQIAQLWKKEEFVISEDQKTNDLVFIQFDKDGMEHFAFLRISLKEQFA 125
 F + + DDL +S IAQLWKE FVISEDQKTNDLVF+QFDKDG FAFLRI+LKEQFA
 10 Sbjct: 61 FMTTIGDDLLETSTVIAQLWKEAFVISEDQKTNDLVFVQFDKDGEPFFAFLRIALKEQFA 120

Query: 126 HVSENQEQPITITQNNLPSAAQTPDEALVVKSSKQYYLIEKRKHNGSFANYFSENLLQ 185
 H+S+N E P T+TQNNLPS QTPDEALV+N S QYYLIEKR+KHNGSFANYFSE+LL+
 Sbjct: 121 HLSDNYEHPFTVTQNNLPSPTQTPDEALVINLKSQYYLIEKRVKHNGSFANYFSEHLLK 180

Query: 186 VQPEQSVKKSIIKMBEQTAQKIAENFNKDDFSQSKMKSATYKNEEEQELSPEKLADQLF 245
 V PEQSVKKSIIKMBEQTAQKIAE+FN+DDF+FSQSKMKS ++K LE + LSPEKLADQLF
 15 Sbjct: 181 VTPEQSVKKSIIKMBEQTAQKIAEHFNQDDFTFSQSKMKSITLKFQLEADDVLSPPEKLADQLF 240

Query: 246 DDNLTLARLTFVDELKDAIPGPVQVSDIDHSRQIKKLENQKLSLSNGIELIVPNNVYQDAE 305
 DDNLTLARLTFVD++KD IP P+++SDI+HSRQIKKLENQKLSLSNGIEL VPN +YQDAE
 20 Sbjct: 241 DDNLTLARLTFVDQVKDVIPEPIKISDIEHSRQIKKLENQKLSLSNGIELTVENAIYQDAE 300

Query: 306 SVEFIQNPDGTYISILIKNIQDIQNK 330
 +VEF+ N DGTYSILIKNI+DI+ K
 25 Sbjct: 301 AVEFLNDDGTYSILIKNIEDIKTK 325

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2125

30 A DNA sequence (GBSx2241) was identified in *S.agalactiae* <SEQ ID 6559> which encodes the amino acid sequence <SEQ ID 6560>. This protein is predicted to be Serine hydroxymethyltransferase (glyA-1). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3876 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD35802 GB: AE001743 serine hydroxymethyltransferase [Thermotoga maritima]
 Identities = 243/416 (58%), Positives = 307/416 (73%), Gaps = 7/416 (1%)

45 Query: 9 KEFDQELWQAIHDEEIRQQNNIELIASENVVSKAVMAAQGSVLTNKYAEGYP SHRYGGT 68
 K+ D E+++ + +E RQ+ +ELIASEN S AV+ GS+LTNKYAEGYP RYGG
 Sbjct: 6 KQVDPEIYEVLVNELEKQYGLIELIASENFASLAVIETMGSMLETNKYAEGYPKKRYGGC 65

50 Query: 69 DCVDVVESLAIERAKTLFNAEFANVQPHSGSQANAAAYMALIEPGDTVLGMDLAAGGHLT 128
 + VD E AIERAK LF A+ANVQPHSGSQAN A Y+AL +PGDT++GM L+ GGHLT
 Sbjct: 66 EWVDRAEERAIERAKRLFGAKFANVQPHSGSQANMAVYLALAQPGDTIMGMSLSHGHLT 125

Query: 129 HGASVSFSGKTYHFVSYSVDPKTEMLDYDNILKIAQETQPKLIVAGASAYSRIIDFEKFR 188
 HGA V+FSGK + V Y V+ +TE +DVD + ++A E +PK+IVAG SAY+RIIDF++FR
 55 Sbjct: 126 HGAPVNFSGKIFKVPYPGVNLETETIDYDEVRRALALEHKPKIIVAGGSAYARIIDFKRFR 185

Query: 189 QIADAVDAYLMVDMAHIAGLVASGHHPSPIPYAHVTTTTTHKTLRGPRGGLILTNDIA 248
 +IAD V AYLMVDMAH AGLVA+G HP+P+ YAHV T+TTHKTLRGPRGGLILTND IA
 60 Sbjct: 186 EIADEVGAYLMVDMAHFAGLVAAGIHPNPLEYAHVVTSTTHKTLRGPRGGLILTNDPEIA 245

Query: 249 KKINSAVFFGLQGGPLEHVIAAKAVALKEALDPSFKIYGEDIKNAQAMAKVFKEDDDFH 308

-2394-

K ++ +FPG+QGGPL HVIAAKAV KEA+ FK Y + ++KNA+ MA+ F++ +
 Sbjct: 246 KAVDKTIFPGIQGGPLMHVIAAKAVCFKEAMTEEFKEYQKQVVKNAKMAEEFQK-RGYR 304
 Query: 309 LISDGTDNHFLVDVTKVIENGKKAQNVLEEVNITLNKNSIPFERLSPFKTSGIRIGTPA 368
 ++S GTD HLFLVD+T GK A+ LE IT+NKN+IP E+ SPF SGIRIGTPA
 Sbjct: 305 IVSGGTDTHLFLVDLTPKDITGKAAEKALESCGITVNKNTIPNEKRSPFVASGIRIGTPA 364
 Query: 369 ITSRGMGVESRRIAELMIKALKN--HENQDVLTEVRQE---IKSLTDAFPLYEN 418
 +T+RGM EE IAE++ L N EN V EVR+E ++ L + FPLY +
 Sbjct: 365 VTTRGMKEBEMEBIAEMIDLVLNSVIDENGTVKPEVREEVSKKVRCLCERFPLYRD 420

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6561> which encodes the amino acid sequence <SEQ ID 6562>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.00 Transmembrane 196 - 212 (196 - 212)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15707 GB:Z99122 serine hydroxymethyltransferase [Bacillus subtilis]
 Identities = 250/407 (61%), Positives = 311/407 (75%), Gaps = 2/407 (0%)
 Query: 14 DKELWDIAHAEERQEHHELIASENMVSKAVMAAQGSLTNKYAEGYPGNRYYGTECV 73
 D+++++AI E ERQ+ IELIASEN VS+AVM AQGSLTNKYAEGYPG RYVGG E V
 Sbjct: 8 DEQVFNAIKNERERQQTKEIASENFVSEAVMAAQGSLTNKYAEGYPGKRYYGCEHV 67
 Query: 74 DIVETLAIERAKLFGAFAFANVQAHSGSQANAAAYMALIEAGDTVLGMDLAAGGHLTHGS 133
 D+VE +A +RAK++FGA NVQ HSG+QAN A Y ++E GDTVLGM+L+ GGHLTHGS
 Sbjct: 68 DVVEDIARDRAKEIFGAHEHVNQPHSGAQANMAVYFTILEQGDVLGMNLSHGHLTHGS 127
 Query: 134 PVNFSGKTYHFVGYSDTIDTEMLNIEAILEQAKAVQPKLIVAGASAYSRSIDFEKFRFRA 193
 PVNFSG Y+FY Y VD +T+ ++Y+ + E+A A +PKLIVAGASAY R+IDF+KFR IA
 Sbjct: 128 PVNFSGVQYNFVEYGVDKETQYIDYDDVREKALAHKPKLIVAGASAYPRITDFKKFREIA 187
 Query: 194 DHVGAYLMVMDMAHIAGLVAAGVHPSPVPYAHIVTSTTHKTLRGPRGGLILTNDALAKKI 253
 D VGAY MVDMAHIAGLVAAG+HP+PVPYA VT+TTHKTLRGPRGG+IL +E KKI
 Sbjct: 188 DEVGAYFMVMDMAHIAGLVAAGLHPNPVPYADFVT+TTHKTLRGPRGMILCREE-FGKKI 246
 Query: 254 NSAVFPGQLGGPLEHVIAAKAVAFKEALDPAFKDYAQAIIDNTAAMAAVFAQDDRFRLIS 313
 + ++FPG+QGGPL HVIAAKAV+F E L FK YAQ +I N +A ++ +L+S
 Sbjct: 247 DKSIFPGIQGGPLMHVIAAKAVSFGEVLQDDFKTYAQNVISNAKRLAEALTKEG-IQLVS 305
 Query: 314 GGTDNHFLVDVTKVIANGKLAQNLLDEVNITLNKNAIPFETLSPFKTSGIRIGCAITS 373
 GGTDNH+ LVD+ + GK+A+++LDE+ IT NKNAIP++ PF TSGIR+G AA+TS
 Sbjct: 306 GGTDNHLILVDLRLSLGLTGKVAEHLVDEIGITSNKNAIPYDPEKPFVTS GIRLGTA AVTS 365
 Query: 374 RGMGVKESQTIARLIKALVNHDQETILEEVRQEVRLTDAFPLYKK 420
 RG + + +I AL NH+ E LEE RQ V LTD FPLYK+
 Sbjct: 366 RGFDDALEEVGAILIALKNHEDEGKLEEARQVAAITDKFFLYKE 412

An alignment of the GAS and GBS proteins is shown below.

Identities = 330/417 (79%), Positives = 368/417 (88%)
 Query: 1 MIFDKDNFKEFDQELWQAIHDEEIRQQNNIELIASENVVSKAVMAAQGSLTNKYAEGYP 60
 MIFDK N ++FD+ELW AIH EE RQ+++IELIASEN+VSKAVMAAQGSLTNKYAEGYP
 Sbjct: 3 MIFDKGNVEDFDKELWDIAHAEERQEHHELIASENMVSKAVMAAQGSLTNKYAEGYP 62
 Query: 61 SHRYYGGTDCVDVVESLAIERAKTLFNAEFANVQPHSGSQANAAAYMALIEPGDVLGMD 120
 +RYYGGT+CVD+VE+LAIERAK LF A FANVQ HSGSQANAAAYMALIE GDTVLGMD

-2395-

Sbjct: 63 GNRYYGGTECVDIVETLAIERAKKLFGAAFANVQAHSQSQANAAAYMALIEAGDTVLGMD 122

Query: 121 LAAGGHLTHGASVSFSGKTYHFVYSVDPKTEMLDYDNILKIAQETQPKLIVAGASAYS 180
LAAGGHLTHG+ V+FSGKTYHFV YSVD TEMPL+Y+ IL+ A+ QPKLIVAGASAYS

5 Sbjct: 123 LAAGGHLTHGSPVNFSGKTYHFVGYSDTDTEMLNIEAILEQAKAVQPKLIVAGASAYS 182

Query: 181 IIDFEKFRQIADAVDAYLMVMDMAHIAGLVASGHHPSPIPYAHVTTTTHKTLRGPRGGLI 240
IDFEKFR IAD V AYLMDVMAHIAGLVA+G HPSP+PYAH+ T+TTHKTLRGPRGGLI

10 Sbjct: 183 SIDFEKFRADHVGAYLMVMDMAHIAGLVAAGVHPSPVPYAHIVTSTTHKTLRGPRGGLI 242

Query: 241 LTND EAI AKKINS AVFPGLQGGPLEHVIAAKAVALKEALDPSFKIYGEDIKNAQAMAKV 300
LTNDEA+AKKINS AVFPGLQGGPLEHVIAAKAVA KEALDP+FK Y + II N AMA V

Sbjct: 243 LTND EAL AKKINS AVFPGLQGGPLEHVIAAKAVAFKEALDPAFKDYAQAIIDNTAAMAAV 302

15 Query: 301 FKEDDDFHLISDGTDNHFLVDVTKVIENCKKAQNVLEEVNITLNKNSIPFERLSPFKTS 360
F +DD F LIS GTDNH+FLVDVTKVI NGK AQN+L+EVNITLNKN+IPFE LSPFKTS

Sbjct: 303 FAQDDRFRLLISGGTDNHVFLVDVTKVIANGKLAQNLLDEVNITLNKNAIPFETLSPFKTS 362

Query: 361 GIRIGTPAITSRGMGVEESRRIABLMIKALKNHENQDVLTEVRQEIKSLTDAFPLYE 417
GIRIG AITSRGMGV+ES+ IA L+IKAL NH+ + +L EVRQE++ LTDAFPLY+

20 Sbjct: 363 GIRIGCAAITSRGMGVKESQTIARLI IKALVNHDQETILEEVRQEVRLTDAFPLYK 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2126

A DNA sequence (GBSx2242) was identified in *Sagalactiae* <SEQ ID 6563> which encodes the amino acid sequence <SEQ ID 6564>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2289(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

A related GBS nucleic acid sequence <SEQ ID 9839> which encodes amino acid sequence <SEQ ID 9840> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35934 GB:AE001752 conserved hypothetical protein [Thermotoga maritima]
Identities = 71/198 (35%), Positives = 114/198 (56%), Gaps = 4/198 (2%)

40 Query: 1 MNDLGQILEDHGAVIMPTETVYGIFAKALSEEAVNHVYELKKRPRDKAMNLNICDFETIL 60
+ + ++L + +I PTETVYGI A A +EEA +++LK+RP D + ++I F+ +

45 Sbjct: 17 LKEAAELLRNGEVIIFPTETVYGIGADAYNEEACKKIFKLKERPADNPLIVHINSFKQLE 76

Query: 61 KYSKNQPTYLKQLYDAFLPGPLTIIL-EASQEVPHWINSGLLSVGFRMPKHPVTLDMIAN 119
+ ++ +L L F PGELT+I + S+++P + + L +V RMP HPV L +I

Sbjct: 77 EIAEGYEPHLDLFL-KKFWPGPLTVIFRKKSEKIPPVVTADLPTVAVRMPAHPVALKLIEL 135

50 Query: 120 HG-PLIGPSANISGCDSCRGVFSEIQQFNHGV-LGIEDDKALTGVSTIIDLSGDRVKIL 177
G P+ PSANISG S + + F +V L I+ G++STI+DL+ ++ +L

Sbjct: 136 FGHPIAAPSANISGRPSATNVKHVIEDFMGKVKLIIDAGDTPFGLESTIVDLTKEKPVLL 195

55 Query: 178 RQGAITQEVLTATIPELI 195
R G + E L PEL+

Sbjct: 196 RGPVPEVERLKLFPPELV 213

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6565> which encodes the amino acid sequence <SEQ ID 6566>. Analysis of this protein sequence reveals the following:

-2396-

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0282(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 127/196 (64%), Positives = 154/196 (77%)

Query: 1 MNDLGQILEDHGAVIMPTETVYGIFAKALSEEAVNHVYELKKRPRDKAMNLCDFETIL 60
 M L I+E A+++PTETVYG+FAKAL E+AVN VY+LK+RPRDKAMNLN+ DF +IL
 Sbjct: 11 MEYLASIIIESGDALVLPETETVYGLFAKALDEKAVNAVYDLKQRPDKAMNLCNVADFNSIL 70

15 Query: 61 KYSKNQPTYLKQLYDAFLPGPLTIILEASQEVPHWINSGLSVGFRMPKHPVTLDMIANH 120
 +SK QP YLK+LY AFLPGPLTIIL+A+ +VP+WINSGL +VGFR+P HP+T +I
 Sbjct: 71 AFSKEQPRYLKKLYQAFLPGPLTIILKANDQVPYWINSGSLSTVGFRLPSPHITAALIQKT 130

20 Query: 121 GPLIGPSANISGCDGRVFEIQQFNHQVLGIEDDKALTGVDSITIDLSGDRVKILRQG 180
 GPLIGPSAN+SG SGRVF I + F+ QV G DD LTG DSTI+DLSG+R ILRQG
 Sbjct: 131 GPLIGPSANLSKASGRVFDHIMQDFDFQVFGYADDFLTGKDDSTILDLSGERAVILRQG 190

25 Query: 181 AITQEVLTATIPELIF 196
 AIT+E L A +PEL F
 Sbjct: 191 AITKEELLANVPELRF 206

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 **Example 2127**

A DNA sequence (GBSx2243) was identified in *S.agalactiae* <SEQ ID 6567> which encodes the amino acid sequence <SEQ ID 6568>. This protein is predicted to be protoporphyrinogen oxidase (hemK). Analysis of this protein sequence reveals the following:

Possible site: 50

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07493 GB:AP001519 protoporphyrinogen oxidase [Bacillus halodurans]
 Identities = 94/236 (39%), Positives = 132/236 (55%), Gaps = 12/236 (5%)

45 Query: 49 DTDQQLMENIFQQLKKHRSP---QYITGKAYFRDLIFFVDERVLI PRPETEELVDLILSE 105
 + D +L + + + L H S Q++ G F F VD+ VLIPRPETEELV +L E
 Sbjct: 46 ELDGELFQRLLEEDLAAHASGVFVQHLIGVESFYGRQFQVDQHVLI PRPETEELVLA VLKE 105

50 Query: 106 -----NKVEDCSVLDIGTGSGAIAISLKKERPSWDVLASDISVSALDLAKENANNCD AEV 160
 K E+ ++LDIGTGSGAIA++L E +V A DIS AL +A +NA A V
 Sbjct: 106 IRRQFKKEEETILDIGTGSGAIAVTLALEEERTINVTAVDISRDALQVAADNARRLGANV 165

55 Query: 161 TFIESDV---FSNISGKFDIIVSNPPYISYNDKDEVGKNVLASEPHSALFADEEGLAIYR 217
 I D+ F +FD+IVSNPPYI +KD + +V EP ALF +GL +YR
 Sbjct: 166 QLIHGDLGEPFLKTGERFDIVSNPPYIPTVEKDTLAVHVRDHEPALALFGGVDGLDVYR 225

60 Query: 218 KIIENSREYL-QPRGKLYFEIGYKQGGDLRLSLKRYFPNNRCRVLKDIFGKDRMVV 272
 +++ + +G + EIG QG D+ L++ +P VL D+ GKDR+V+
 Sbjct: 226 RLMSQLPALTKKEKGMVALEIGAGQGMVLEKLMQTAYPKAAVDVLYDLNGKDRIVL 281

-2397-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6569> which encodes the amino acid sequence <SEQ ID 6570>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 48
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.4324(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

   Identities = 174/274 (63%), Positives = 207/274 (75%)

15  Query: 1  MNYAQLIKHYGQLEACGEEVENFIYVLKDLKQWSTTDYLLNQSSVSDTDQQLMENIFQ 60
     Sbjct: 1  MNYA LI+ Y  LE  E+ EN  YV +++K+WS+ D L++QN +V+  D  L+E+IF
     MNYA TLIRTYEDKLEQIDEDRENLAYVFREIKWSSLDMLIHQNAVTPEDAVLLEHIFC 60

20  Query: 61  QLKXHRSPQYITGKAYFRDLIFFVDERVLIPRPETEELVDLILSENKVEDCSVLDIGTGS 120
     Sbjct: 61  L +H SPQYITG AYFRDL  VD+RVLIPRPETEELVD+IL+EN  +VLDIGTGS
     SLSQHLSPQYITGNAYFRDLKLAVDKRVLIPRPETEELVDMILAENLDAPLNVLVDIGTGS 120

25  Query: 121 GAIAISLKKERP+WDVLASDISVSALDLAKENANNCDAEVTFIESDVFSNISGKFDIIVS 180
     Sbjct: 121 GAIAISLKKERP+W V ASDIS +ALDLAK NA+  ++TFIESDVFS IS  FDIIVS
     GAIAISLKKERP+WDVLASDISVSALDLAKENANNCDAEVTFIESDVFSNISGKFDIIVS 180

30  Query: 181 NPPYISYNDKDEVGKNVLASEPHSALFADEEGLAIYRKIIENSREYLQPRGKLYFEIGYK 240
     Sbjct: 181 NPPYISY DK+EV  NVL SEPH ALFA E G AIYRKIIE +  YL  GKLYFEIGYK
     NPPYISYNDKDEVGKNVLASEPHSALFADEEGLAIYRKIIENSREYLQPRGKLYFEIGYK 240

   Query: 241 QGDDLRSLKRYFPNNRCRVLKDIFGKDRMVVD 274
     Sbjct: 241 Q + ++ +L+ YFP  R + DIFGK+RMVV+D
     QGDDLRSLKRYFPNNRCRVLKDIFGKDRMVVD 274
     QAEGIKDMLQAYFPQRHRAVTDIFGKERMVVVD 274

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2128

A DNA sequence (GBSx2244) was identified in *S.galactiae* <SEQ ID 6571> which encodes the amino acid sequence <SEQ ID 6572>. This protein is predicted to be peptide chain release factor RF-1 (prfA).

Analysis of this protein sequence reveals the following:

```

   Possible site: 28
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
45  bacterial cytoplasm --- Certainty=0.3446(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50  >GP:CAB15718 GB:Z99122 peptide chain release factor 1 [Bacillus subtilis]
     Identities = 211/351 (60%), Positives = 280/351 (79%), Gaps = 1/351 (0%)

   Query: 5  DQLQAVEDRYEELGELLSDPDVVS DTKRFMELSREEASTRET VTYREYKQVIQNISDAE 64
     Sbjct: 3  D+L++++E+RYE+L ELLSDP+VV+D K+  E S+E++ +ETV YR+Y+  + ++DA+
55  DRLKSIEERYEKLNELLSDP EVVNDPKKLREYSKEQSDIQETVDVYRQYRDASEQLADAK 62

   Query: 65  EMIKDASGDABEEMAKEELKESKAAKEEYERL KILL LKDPND DKN ILEIRGAAGGD 124
     M+++  DAE+ +M KEE+ E +  E  ERLK+LL+PKDPND DKN +I+EIRGAAGG+

```

-2398-

5 Sbjct: 63 AMLEEKL-DAEMRDMVKEEISELQKETETLSERLKVLLIPKDPNDCKNVIMEIRGAAGGE 121
 Query: 125 EAALFAGDLLTMYQKYAETQGWRFVMESSVNGVGGIKEVVAMVSGQSVYSLKYESGAH 184
 EAALFAG+L MY +YAE QGW+ EVME++V G GG KE++ M++G YSKLYE+GAH
 10 Sbjct: 122 EAALFAGNLYRMYRYAELQGWKTEVMEANVTGTGGYKEIIFMITGSGAYSCLKYENGAAH 181
 Query: 185 RVQRVPVTESQGRVHTSTATVLMPEVEEVEYEIDQKDLRVDIYHASGAGGQNVNKVATA 244
 RVQRVP TES GR+HTSTATV +PE EEVE +I +KD+RVD + +SG GGQ+VN +A
 15 Sbjct: 182 RVQRVPVTESGGRIHTSTATVACLPEAEVEVDIHEKDIRVDTFASSGPGGQSVNNTMSA 241
 Query: 245 VRMVHIPTGIKVMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVGTGDR 304
 VR+ H+PTG+ V Q+E++Q KN++KAMK++RAR+ D F Q AQ E D RKS VG+GDR
 Sbjct: 242 VRLTHLPTGVVSCQDEKSQLKNKEKAMKVLARARYDKFQQEQAQAYDQTRKSAVGSQDR 301
 20 Query: 305 SERIRTYNFPQNRVTDHRIGLTQKLDITLSGKMDEVIDALVMDQTQKLE 355
 SERIRTYNFPQNRVTDHRIGLT+QKLD IL GK+DEV++AL++ DQ KL+
 Sbjct: 302 SERIRTYNFPQNRVTDHRIGLTQKLDQILEGKLDEVVEALIVEDQASKLQ 352

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6573> which encodes the amino acid
 sequence <SEQ ID 6574>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3446(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 349/358 (97%), Positives = 354/358 (98%)
 Query: 1 MNIYDQLQAVEDRYEELGELLSDPDVVS DTKRFMELSREEASTRETVTAYREYKQVIQNI 60
 MNIYDQLQAVEDRYEELGELLSDPDVVS DTKRFMELSREE +TRETVTAYREYKQVIQ I
 35 Sbjct: 1 MNIYDQLQAVEDRYEELGELLSDPDVVS DTKRFMELSREETNTRETVTAYREYKQVIQTI 60
 Query: 61 SDAEEMIKDASGDAELEEMAKEELKESKAAKEEYERLKIILLPKDPNDCKNIILEIRGA 120
 SDAEEMIKDASGD ELEEMAKEELKESKAAKEEYEE+LKILLPKDPNDCKNIILEIRGA
 Sbjct: 61 SDAEEMIKDASGDPELEEMAKEELKESKAAKEEYEEKLKIILLPKDPNDCKNIILEIRGA 120
 40 Query: 121 AGGDEAALFAGDLLTMYQKYAETQGWRFVMESSVNGVGGIKEVVAMVSGQSVYSLKYE 180
 AGGDEAALFAGDLLTMYQKYAETQGWRFVMESSVNGVGGIKEVVAMVSGQSVYSLKYE
 Sbjct: 121 AGGDEAALFAGDLLTMYQKYAETQGWRFVMESSVNGVGGIKEVVAMVSGQSVYSLKYE 180
 45 Query: 181 SGAHRVQRVPVTESQGRVHTSTATVLMPEVEEVEYEIDQKDLRVDIYHASGAGGQNVNK 240
 SGAHRVQRVPVTESQGRVHTSTATVLMPEVEEVEY+ID KDLRVDIYHASGAGGQNVNK
 Sbjct: 181 SGAHRVQRVPVTESQGRVHTSTATVLMPEVEEVEYDIDPKDLRVDIYHASGAGGQNVNK 240
 Query: 241 VATAVRMVHIPTGIKVMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVG 300
 VATAVRMVHIPTGIKVMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVG
 50 Sbjct: 241 VATAVRMVHIPTGIKVMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVG 300
 Query: 301 TGDRSERIRTYNFPQNRVTDHRIGLTQKLDITLSGKMDEVIDALVMDQTQKLEALN 358
 TGDRSERIRTYNFPQNRVTDHRIGLTQKLDITLSGKMDEVIDALVMDQT+KLE+LN
 55 Sbjct: 301 TGDRSERIRTYNFPQNRVTDHRIGLTQKLDITLSGKMDEVIDALVMDQTKKLES LN 358

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

-2399-

Example 2129

A DNA sequence (GBSx2245) was identified in *S.agalactiae* <SEQ ID 6575> which encodes the amino acid sequence <SEQ ID 6576>. This protein is predicted to be thymidine kinase (tdk). Analysis of this protein sequence reveals the following:

```

5   Possible site: 39
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2244 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9841> which encodes amino acid sequence <SEQ ID 9842> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB02289 GB:L40415 thymidine kinase [Streptococcus gordonii]
Identities = 158/189 (83%), Positives = 175/189 (91%)

Query: 1  MAQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDEFGVSSRIGMRREAV 60
20      MAQLYYKYGTMSGKTIEILKVAHNYEEQGK VVIMTSA+DTRD G VSSRIGM+R+A+
Sbjct: 1  MAQLYYKYGTMSGKTIEILKVAHNYEEQGKGVVIMTSAVDTRDGVGVSSRIGMKRQAM 60

Query: 61  PISDDMDIFSITQNLFPQKPYCVLIDECQFLSKKNVYDLARVVDDLDPVMAFGLKNDFFQN 120
          I DD DI YI+NLP+KPYC+LIDE QFL + +VYDLARVVD+LDVPVMAFGLKNDFF+N
25      Sbjct: 61  AIEDDTDILGYIKNLEKPYCILIDEAQFLKRHHVYDLARVVDELDVPVMAFGLKNDFFRN 120

Query: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180
          LFEFGSKHLLLLADKI+EIKTICQYCS+KATMVLRT++GKPVY+G+QIQIGGNETYIPVC
30      Sbjct: 121 ELFEFGSKHLLLLADKIEEIKTICQYCSRKATMVLRTDHGKPVYDGEQIQIGGNETYIPVC 180

Query: 181 RKHYFNPDI 189
          RKHYF PDI
35      Sbjct: 181 RKHYFKPDI 189

```

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6577> which encodes the amino acid sequence <SEQ ID 6578>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

40      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2244 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 An alignment of the GAS and GBS proteins is shown below.

```

Identities = 174/189 (92%), Positives = 184/189 (97%)

Query: 1  MAQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDEFGVSSRIGMRREAV 60
50      +AQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRD FG+VSSRIGMRREA+
Sbjct: 1  LAQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDGFVGIVSSRIGMRREAI 60

Query: 61  PISDDMDIFSITQNLFPQKPYCVLIDECQFLSKKNVYDLARVVDDLDPVMAFGLKNDFFQN 120
          PIS+DNDIF++I L +KPYCVLIDE QFLSK+NVYDLARVVD+L+VPVMAFGLKNDFFQN
55      Sbjct: 61  PISNDMDIFTFIAQLEKPYCVLIDESQFLSKQNVYDLARVVDELNVVMAFGLKNDFFQN 120

Query: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180
          NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC
          Sbjct: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180

```

-2400-

Query: 181 RKHYFNPD1 189
 RKHYFNPD1
 Sbjct: 181 RKHYFNPD1 189

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2130

A DNA sequence (GBSx2246) was identified in *S.agalactiae* <SEQ ID 6579> which encodes the amino acid sequence <SEQ ID 6580>. Analysis of this protein sequence reveals the following:

10 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3995(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAA26046 GB:M95650 4-oxalocrotonate tautomerase [Plasmid pWW0]
 Identities = 27/60 (45%), Positives = 36/60 (60%)

Query: 1 MPFVKIDLFEGRSQEQKNELAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGELKKK 60
 MP +I + EGRS EQK L REV+E +SR AP ++ V I +M +G + GEL K
 25 Sbjct: 1 MPFIAQHILEGRSDEQKETLIREVSEAIRSLDAPLTSVRVIITEMAKGHFGIGGELASK 60

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6581> which encodes the amino acid sequence <SEQ ID 6582>. Analysis of this protein sequence reveals the following:

30 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.4128(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 56/60 (93%), Positives = 59/60 (98%)

40 Query: 1 MPFVKIDLFEGRSQEQKNELAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGELKKK 60
 MPFV IDLFEGRSQEQKN+LAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGE+K+K
 Sbjct: 1 MPFVTIDLFEGRSQEQKNQLAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGEMKQK 60

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 2131

A DNA sequence (GBSx2247) was identified in *S.agalactiae* <SEQ ID 6583> which encodes the amino acid sequence <SEQ ID 6584>. Analysis of this protein sequence reveals the following:

50 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2154(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2401-

A related GBS nucleic acid sequence <SEQ ID 9843> which encodes amino acid sequence <SEQ ID 9844> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAC65759 GB:AE001250 conserved hypothetical protein [Treponema
    pallidum]
    Identities = 103/317 (32%), Positives = 163/317 (50%), Gaps = 15/317 (4%)

10 Query: 7  QLSHSLRLMGTITIDIQINSKNAQKQIR---EVIELLELYKNRFSANDFNSELMMAINNA 62
    + S + ++GT +++ SK ++ EV LL+ + SAN +S L A+N A
    Sbjct: 31 EYSRAELVIGTLCRVRVYSKRPAAEVHAALAEVFTLLQQQEMVLSANRDDSALAALNAQA 90

    Query: 63 GIKPIQVHPDLFELITIGKEHSLARPSNLNIAIGPLVQTRIGFSDAKLPSPSEISEAMI 122
    G P+ V L+ L+ + N A+G V+ W IGF A +P P + EA+
15 Sbjct: 91 GSAPVVVDRSLYALLERALLFAEKSGGAFNPALGAXVKLWNIGFDRAAVPDPDALKEALT 150

    Query: 123 LSDPTHILLDSN-----KQSVFLNQIGMKIDLALAGYIADKIMTYLKNEMIDSAIINL 177
    D + L + +V L Q GM++DLGA+AKG++ADKI+ L +DSA+++L
20 Sbjct: 151 RCDFRQVHLRAGSVSGAPHTVQLAQAGMQLDLGAIAGFLADKIVQLLTAHALDSALVDL 210

    Query: 178 GGNV---LVHGDNPNRSEGY--WVIGIQHPKKRGKNIGTVKIKNQSVVTSGTYERRLI 231
    GGN+ L +GD + + W +GI+ P K V +++ SVVTSG YER
    Sbjct: 211 GGNIFALGLKYGDVRSAAQRLAWNNGIRDPHGTGQKPALVVSVRDCSVVTSGAYERFFE 270

25 Query: 232 IDDKKEYHHIFDRQTGYPIQTEMASISIVSKQSDCEIWTTRLFGLSIKEALDILNAVSYI 291
    D YHHI D TG+P T++ S+SI + +S D + T F L +++ +L +
    Sbjct: 271 RDGVRVYHHIIDPVTGFPAHTDVDSVSIFAPRSTDADALATACFVLGYEKSCALLREBFGV 330

30 Query: 292 EGIITTKDDRIYLSDDL 308
    + + I D R+ S G+
    Sbjct: 331 DALFIFPDKRVRASAGI 347

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6585> which encodes the amino acid sequence <SEQ ID 6586>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 52
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.1020(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

45 Identities = 182/310 (58%), Positives = 232/310 (74%)

    Query: 8  LSHSLRLMGTITIDIQINSKNAQKQIREVIELLELYKNRFSANDFNSELMMAINNAGIKPI 67
    ++ L+LMGT IDIQI S A +Q+ VI+LL YKNRFSAND NSELMMAIN AG+KP+
    Sbjct: 3  VTQQLKLMGTVIDIQIESDKACQQLSRVIDLLTYKNRFSANDSNSELMMAINQAAGVKPV 62

50 Query: 68  QVHPDLFELITIGKEHSLARPSNLNIAIGPLVQTRIGFSDAKLPSPSEISEAMILSDPT 127
    VH DLF LI IGK HSL+ PSNLNIAIGPLVQ WRIGF DA++PS + IS+ + L+DP
    Sbjct: 63 SVHSDLFNLIQIGKAHSLSTPSNLNIAIGPLVQAWRIGFEDARVPSHNLISQQLALTDPR 122

55 Query: 128 HILLDSNKQSVFLNQIGMKIDLALAGYIADKIMTYLKNEMIDSAIINLGGNVLVHGDN 187
    +L+D KQ+VFL Q+GM +DLGALAGYI DKIM YL + IDSA+INLGGNV VH G N
    Sbjct: 123 QVLIDDKKQTVFLQVGMALDLGALAGYITDKIMAYLIEDGIDSALINLGGNVRVHGPV 182

    Query: 188 PNRSEGYWVIGIQHPKKRGKNIGTVKIKNQSVVTSGTYERRLIIDDKKEYHHIFDRQTGY 247
    P + + IGIQ P KRG+++G +K+ N SVVTSG YER+ K+YHHI DRQTGY
60 Sbjct: 183 PKSPDKTFRIGIQKPDARKGQHLGVKVNNSVVTSGIYERQFTSKGQYHHILDRTQGY 242

    Query: 248 PIQTEMASISIVSKQSDCEIWTTRLFGLSIKEALDILNAVSYIEGIITTKDDRIYLSDLG 307

```

-2402-

PI+T+M S++I++ S C+IWTTTLFGL + +LN IEG+++T+ + +S+G
 Sbjct: 243 PIETDMLSLTIMAPSSFYCDIWTTRLFGLDSSMIITLNTFDNIEGLLVTRKHHVLSNG 302
 Query: 308 LKHFFQLFYH 317
 L+H+FQ +YH
 Sbjct: 303 LRHYFQPYH 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 2132

A DNA sequence (GBSx2248) was identified in *S.agalactiae* <SEQ ID 6587> which encodes the amino acid sequence <SEQ ID 6588>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0966(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG18632 GB:AY007504 unknown [Streptococcus mitis]
 Identities = 92/160 (57%), Positives = 119/160 (73%), Gaps = 1/160 (0%)
 Query: 1 MKLIGIVGTNSNKSTNRQLLQYMQQHFAADKAEIELIEVKDLPLFNKPADKNVPQVILDIA 60
 MKL+ IVGTNSN+STNR+LL++MQ+HF+DKA+IE++E+K LP FN+P D+ P + +
 Sbjct: 1 MKLVAIVGTNSNRSTNRKLLKFMQKHFSKADIEVLEIKQLPAFNEPEDEQAPAEVQAFS 60
 Query: 61 AKIETDGVIIIGTPEYDHSIPSALMSVLAWLSYGIYPLLNKPMITGASYGTLGSSRAQL 120
 KI DGVII TPEYDH+IP+ L S L W++Y L+NKP MI GAS G LG+SRAQ
 Sbjct: 61 EKILAADGVIIISTPEYDHTIPAPLASALEWLIAYTSRALINKPTMIVGASLGLLGTSRAQA 120
 Query: 121 QLRQILNAPELKASVLP-DEFLSHSLQAFDKDGNLHDIE 159
 LRQIL+APELKA V+P EF L HS Q D + +L++ E
 Sbjct: 121 HLRQILDAPELKARVMPGTEFFLGHSEQLDDECHLNPE 160

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6589> which encodes the amino acid sequence <SEQ ID 6590>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB62679 GB:AL133422 putative secreted protein. [Streptomyces coelicolor A3(2)]
 Identities = 68/192 (35%), Positives = 94/192 (48%), Gaps = 25/192 (13%)
 Query: 4 ILFIVGSLREGSFNHQLAAQAQK-ALEHQAVVSILNWKDVPVLNQDIEANAPLPVVDA-- 60
 IL +VGS LR GS N QLA A + A E V + ++P N+DI+ +P A
 Sbjct: 5 ILALVGS LRAGSHNRQLAEAAVRFAPEGAEVQLFEGLAEIPFYNEIDIDVEGSVPAAAAKL 64
 Query: 61 RQAVQSADAIWIFTPVYNFSIPGSVKNLLDWSLRALDLSDPGSAIGGKVVTVSSVANG 120
 R+A Q A A +F+P YN +IP +KN +DWLSR P G A GK V V A G
 Sbjct: 65 REAAQGAQAFLLFSPEYNGTIPAVLKNAIDWLSR-----PYGAGFTGKPVAVVGTAFG 118

-2403-

Query: 121 GHDQVFDQFKA-----LLPFIRTSVAGEFTK-ATVNP--DAWGTGRLEISKETKA 167
 + V+ Q +A ++ I+ S+ G T+ A +P DA +L E A
 Sbjet: 119 QYGGVWAQDEARKAVGIAGGKVIEDIKLSIPGSVTRFAETHPADDAEVAAQL---TEVVA 175

Query: 168 NLLSQAEALLAA 179
 L A+ +AA
 Sbjet: 176 RLHGHADAIAIA 187

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 28/90 (31%), Positives = 49/90 (54%)

Query: 3 LIGIVGTNSNKSINRQLLQYMQQHFADKAEIELIEVKDLPLFNKPADKNVPQVILDIAAK 62
 ++ IVG+ S N QL Q+ +A + + KD+P+ N+ + N P ++D

15 Sbjet: 4 ILFIVGSLREGSFNHQLAAQAQKALEHQAVVSYLNWKDVPVLNQDIEANAPLPVVDARQA 63

Query: 63 IEETDGVIIIGTPEYDHSIPSALMSVLWLS 92
 ++ D + I TP Y+ SIP ++ ++L WLS

20 Sbjet: 64 VQSADAIWIFTPVYNFSIPGSVKNLLDWLS 93

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2133

25 A DNA sequence (GBSx2249) was identified in *S.agalactiae* <SEQ ID 6591> which encodes the amino acid sequence <SEQ ID 6592>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1160(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2134

40 A DNA sequence (GBSx2250) was identified in *S.agalactiae* <SEQ ID 6593> which encodes the amino acid sequence <SEQ ID 6594>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2132(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAG18632 GB:AY007504 unknown [Streptococcus mitis]
 Identities = 80/162 (49%), Positives = 112/162 (68%)

Query: 1 MKFVGIVGNSAEQSYNRMILLEFIRKNFKTKFELEVLEIDDIPMFNQDQNWEEFQLRLLN 60

-2404-

```

      MK V IVG+N+ +S NR LL+F++K+F K ++EVLEI +P FN+ ++ + +++ +
Sbjct: 1 MKLVAVGTNSNRSTNRKLLKFMQKHFSKADIEVLEIKQLPAFNEPEDEQAPAEVQAFS 60

Query: 61 NKITRADGVIIATPEHNHTITAALKSVLEWLSFAVHPLENKPVMIVGASYDQGTSTRAQI 120
      KI ADGVII+TPE++HTI'A L S LEW+++ L NKP MIVGAS GTSRAQ
Sbjct: 61 EKILAADGVIIISTPEYDHTIPAPLASALEWIAYTSTRALINKPTMIVGASLGLLGTSTRAQA 120

Query: 121 HLRKILDAPGVNAYTLPGNEFLLGKAKEAFDDNGNIINPGTV 162
      HLR+ILDAP + A +PG EF LG +++ DD ++ NP V
Sbjct: 121 HLRQILDAPELKARVMGTEFFLGHSEQVLDDECHLNPEKV 162

```

There is also homology to SEQ ID 6596.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 2135

A DNA sequence (GBSx2251) was identified in *S.agalactiae* <SEQ ID 6597> which encodes the amino acid sequence <SEQ ID 6598>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence
20  INTEGRAL    Likelihood = -7.32    Transmembrane    13 - 29 ( 11 - 29)

----- Final Results -----
      bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2136

A DNA sequence (GBSx2252) was identified in *S.agalactiae* <SEQ ID 6599> which encodes the amino acid sequence <SEQ ID 6600>. This protein is predicted to be potential nitrite transporter. Analysis of this protein sequence reveals the following:

```

35  Possible site: 42
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.92    Transmembrane    61 - 77 ( 54 - 82)
      INTEGRAL    Likelihood = -5.57    Transmembrane    106 - 122 ( 103 - 126)
      INTEGRAL    Likelihood = -5.15    Transmembrane    160 - 176 ( 159 - 177)
40  INTEGRAL    Likelihood = -4.09    Transmembrane    180 - 196 ( 179 - 199)
      INTEGRAL    Likelihood = -1.01    Transmembrane    233 - 249 ( 233 - 249)

----- Final Results -----
      bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15832 GB:Z99123 alternate gene name: ipa-48r~similar to
50  nitrite transporter [Bacillus subtilis]
      Identities = 82/253 (32%), Positives = 119/253 (46%), Gaps = 10/253 (3%)
Query: 6  EKIAYNCAKKEALYKESLGRYALRSMLAGAYLTMSTAAGIVAADTICK-ISPALSGFVF- 63

```

-2405-

+K+ KK+ ++ S RY LRS+LA ++ GI AA G A S F F
 Sbjct: 7 QKVEQYALKKKQNI FASSKIRYVLR SILASIFIGF----GITAASKTGSYFFMADSPFAFP 62
 Query: 64 --AFIFSGLIYVLIFNGELATSNM LYLTAGAYNKNISWKKAMTILYCTFFNLVGACIL 121
 5 A F ++ + G+L T N Y T A K ISW+ + + + NL+GA +
 Sbjct: 63 AAATVFGAAILMIAYGGGDLFTGNTFYFTYTTALRKKISWRD TLVLMSSYAGNLIGAILF 122
 Query: 122 AWLFNQSYSFQHLTND SFLGHVVAKKLGKPPSSGAFLEGIIANMFVNLA ILAYMLLKEESA 181
 10 A L + + F+ + SFL H+ K+ P+S F G++ N V LA M LK E A
 Sbjct: 123 AILISATGLFEEPSVHSFLIHLAEHKMEPPASELFFRGMLCNWL VCLAFFIPMSLKGE GA 182
 Query: 182 KMTVILSAIFMFVFLSNEHLIANFASFMLA AFSHIEHIKGF TLLNIIRQWTLVFFGNWIG 241
 15 K+ ++ +F F EH IAN +F ++ IEH TL+ +R V GN
 Sbjct: 183 KLFTMMLFVFCFFISGF EHSIANMCTFAISLL--IEHPDTVTLMGAVRNLI PVTLGNLTA 240
 Query: 242 GGVFIGLAYAWLN 254
 20 G V +G Y LN
 Sbjct: 241 GIVMMGWMYYTLN 253

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6601> which encodes the amino acid sequence <SEQ ID 6602>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -9.77 Transmembrane 142 - 158 (139 - 171)
 INTEGRAL Likelihood = -9.34 Transmembrane 95 - 111 (89 - 119)
 INTEGRAL Likelihood = -2.02 Transmembrane 61 - 77 (61 - 79)
 INTEGRAL Likelihood = -1.12 Transmembrane 261 - 277 (261 - 279)
 INTEGRAL Likelihood = -0.53 Transmembrane 191 - 207 (191 - 207)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:AAB80864 GB:U93874 formate dehydrogenase [Bacillus subtilis]
 Identities = 133/258 (51%), Positives = 181/258 (69%)
 40 Query: 36 KTPEQILEATIHIGEHKVTKTFLAKSILGFIGGAMISLG YLLYVRIAASGLETFGAFSSI 95
 + P++I EA I G K+ + +LGF+GGA I+LGYLL +R+ + +G+ SS+
 Sbjct: 4 RKPDEIAEAAIEAGMKKIKLPLPSLLVLGLGAFIALGYLLDIRVIGDLPKEWGSLSL 63
 Query: 96 VGACAFPIGLIIILMAGGELITGNMMAVSAALLAKKIKFSELAKNWLIITL FNVIGAVFV 155
 +GA FP+GLI++++AG ELITGNMM+V+ AL ++KI ELA NW I+T+ N+IGA+FV
 45 Sbjct: 64 IGAAVFPVGLILVVLGAEELITGNMMSVAMALFSRKISVKELAINWGI VTIMNLIGALFV 123
 Query: 156 AFVFGHFLGLTSAGIFKEEVIEVAHAKIAASPLQALVSGIGCNWFVGLALWLCY GANDAA 215
 A+ FGH +GLT G + E+ I VA K+ S + L+S IGCNW V LA+WL +GA DAA
 50 Sbjct: 124 AYFFGHVLGLTETGPYLEKTIAVAQGLDMSFGKVLISAIGCNWL VCLAVWLSFGAQDAA 183
 Query: 216 GKFLGTWFPVMTFVALGFGHVSANAFVIPAAIFEGGATWLD FVTNFIFVYSGNIIGGAIF 275
 GK LG WFP+M FVA+GFQH VAN FVIPAAIF G TW F+ N I + GN+IGGA+F
 Sbjct: 184 GKILGIWFPIMAFVAIGFQHVANMFVIPAAIFAGSFTWGFIGNIIPAFIGNVIGGAVF 243
 55 Query: 276 VSFLYFKVYVYHPQKSKTQ 293
 V +YF Y+ +S+ +
 Sbjct: 244 VGLIYFIAYHKDRSRKE 261

An alignment of the GAS and GBS proteins is shown below.

60 Identities = 69/240 (28%), Positives = 101/240 (41%), Gaps = 18/240 (7%)
 Query: 15 KEALYKESLGRYALRSMLAGAYLTMTSTAAGIVAADTIGKISPALSGFVF AFIFSFGLIYV 74
 K L K LG + G L + AA +T G A S V A F GLI +
 Sbjct: 55 KTFLAKSILGFIGGAMISLG YLLYVRIAAS--GLETFG----AFSSIVGACAFPIGLIII 108

-2406-

Query: 75 LIFNGELATSNMLYLTAGAYKNKNISWKKAMTILLYCTFFNLVGACILAWLIFNQSYSFQHL 134
 L+ GEL T NM+ ++A K I + + + T FN++GA +A++F F L
 Sbjct: 109 LMAGELITGNMMAVSAALLAKKIKFSELAKNWLIITLTFNVIGAVFVAFVFGH---FLGL 165

Query: 135 TNDSPFLGHVVAK----KLGKPSGGAFLEGIIANMFVNLAAILAYMLLKEESAKMTVILSAI 190
 T+ V + K+ A + GI N FV LA+ + + K +
 Sbjct: 166 TSAGIFKEEVIEVAHAKIAASPLQALVSGIGCNWVGLALWLCYGANDAAGKFLGTWFPV 225

Query: 191 FMFVPLSNEHLIANFASFMLAASFHIEHIKGFLLNIIRQWTLVFFGNWIGGGVFIGLAY 250
 FV L +H +AN A F G T L+ + + V+ GN IGG +F+ Y
 Sbjct: 226 MTFVALGFGHVSANAFVIPAAIFE-----GGATWLDFTVTFIFVYSGNIIGGAIFVVSFLY 280

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2137

A DNA sequence (GBSx2253) was identified in *S.agalactiae* <SEQ ID 6603> which encodes the amino acid sequence <SEQ ID 6604>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1342(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2138

A DNA sequence (GBSx2254) was identified in *S.agalactiae* <SEQ ID 6605> which encodes the amino acid sequence <SEQ ID 6606>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.22 Transmembrane 44 - 60 (44 - 60)

----- Final Results -----
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2407-

Example 2139

A DNA sequence (GBSx2255) was identified in *S.agalactiae* <SEQ ID 6607> which encodes the amino acid sequence <SEQ ID 6608>. This protein is predicted to be xanthine permease (pbuX). Analysis of this protein sequence reveals the following:

```

5   Possible site: 23
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.91    Transmembrane  160 - 176 ( 156 - 188)
      INTEGRAL    Likelihood = -6.48    Transmembrane  184 - 200 ( 179 - 211)
      INTEGRAL    Likelihood = -6.21    Transmembrane  101 - 117 (  96 - 121)
10  INTEGRAL    Likelihood = -4.04    Transmembrane  309 - 325 ( 306 - 332)
      INTEGRAL    Likelihood = -3.98    Transmembrane  334 - 350 ( 331 - 353)
      INTEGRAL    Likelihood = -3.88    Transmembrane  400 - 416 ( 396 - 420)
      INTEGRAL    Likelihood = -3.45    Transmembrane   19 -  35 (  18 -  38)
      INTEGRAL    Likelihood = -2.81    Transmembrane  127 - 143 ( 127 - 146)
15  INTEGRAL    Likelihood = -2.71    Transmembrane  228 - 244 ( 227 - 249)
      INTEGRAL    Likelihood = -2.02    Transmembrane   47 -  63 (  47 -  63)
      INTEGRAL    Likelihood = -1.97    Transmembrane   75 -  91 (  73 -  92)
      INTEGRAL    Likelihood = -0.85    Transmembrane  368 - 384 ( 368 - 384)

20  ----- Final Results -----
      bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25  The protein has homology with the following sequences in the GENPEPT database.
      >GP:CAB14123 GB:Z99115 xanthine permease [Bacillus subtilis]
      Identities = 213/412 (51%), Positives = 292/412 (70%), Gaps = 5/412 (1%)

30  Query: 14  LGLQHLLAMYAGSILVPIMIASALGYNAKQLTYLIATDIFMCGIATLLQLRLSKHFGVGL 73
      LG+QH+LAMYAG+I+VP+++ A+G +QLTYL++ DIFMCG+ATLLQ+ ++ FG+GL
      Sbjct: 11  LGIQHVLAMYAGAIVVPLIVGKAMGLTVBQLTYLVSIDIFMCGVATLLQVWSNRFFGIGL 70

      Query: 74  PVVLGCAFSQSVAPLSIIQAQQSGYMGFALIASGIYVVLVAGIFSKVANFFPPPIVTGSVI 133
      PVVLGC F +V+P+ IG++ G ++G++IASGI V+L++ F K+ +FFPP+VTGSV+
35  Sbjct: 71  PVVLGCTFTAVSPMIAIGSEYGVSTVYGSIIASGILVILISFFFGKLVSFPPPVVTGSVV 130

      Query: 134 TTIGTLTLPVAMGNMGD---NAKEPSLQSLTSLVTIGVLLINIFAKGFLKSISILIGL 190
      T IG+TL+PVAM NM +A L +L L+ + +++L+ F KGF+KS+SILIG+
40  Sbjct: 131 TIIGTITLMPVAMNNMAGGEGSADFGDLNLAALFTVLSIIIVLLYRFTKGFISVSILIGI 190

      Query: 191 ISGTILAAFMGLVDASVVADAPLVHIPKPFYFGAPREFEFTSILMMCIATVSMVESTGVY 250
      + GT +A FMG V V+DA +V + +PFYFGAP F I+ M I+A VS+VESTGVY
      Sbjct: 191 LIGTFIAYFMGKVQFDNVSDAAVQMIQPFYFGAPSFHAAPITMSIVAIVSLVESTGVY 250

45  Query: 251 LALSDITNDKLDKRLRNGYRSEGLAVLLGGLFNTFFPYTGFSQNVGLVQISGIRTRKPIY 310
      AL D+TN +L L GYR+EGLAVLLGG+FN FPYT FSQNVGLVQ++GI+ I
      Sbjct: 251 FALGDLTNRRLTEIDLSKGYRAEGLAVLLGGIFNAFPYTAFSQNVGLVQLTGIIKKNNAVIV 310

      Query: 311 FTALFLVLILGLLPKFGAMAQMIPSPVLGGAMLVLFMGMVALQGMKMLNQVDFEHNEHNFII 370
      T + L+ GL PK A +IPS VLGGAM+ +FGMV G+KML+++DF E N +I
50  Sbjct: 311 VTGVILMAFGLFPKIAAFTTIIPSAVLGGAMVAMFGMVIAYGIKMLSRIDFAKQE-NLLI 369

      Query: 371 AAVSIAAGVGFGNGT-NLFISLPNTLQMFLTNGIVISTLTAVVLNIIINGLPK 421
      A S+ G+G ++F LP+ L + TNGIV + TAVVLNI+ N K
55  Sbjct: 370 VACSVGLGLGVTVVPDIFKQLPSALTLLTNGIVAGSFTAVVLNIVYNVFSK 421

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6609> which encodes the amino acid sequence <SEQ ID 6610>. Analysis of this protein sequence reveals the following:

```

60  Possible site: 29
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.32    Transmembrane  160 - 176 ( 158 - 181)
      INTEGRAL    Likelihood = -6.37    Transmembrane  103 - 119 (  98 - 124)

```

-2408-

INTEGRAL Likelihood = -5.84 Transmembrane 130 - 146 (126 - 152)
 INTEGRAL Likelihood = -5.68 Transmembrane 187 - 203 (182 - 207)
 INTEGRAL Likelihood = -3.98 Transmembrane 337 - 353 (334 - 356)
 INTEGRAL Likelihood = -3.82 Transmembrane 232 - 248 (225 - 252)
 5 INTEGRAL Likelihood = -3.35 Transmembrane 403 - 419 (399 - 421)
 INTEGRAL Likelihood = -2.50 Transmembrane 22 - 38 (21 - 41)
 INTEGRAL Likelihood = -2.07 Transmembrane 312 - 328 (312 - 328)
 INTEGRAL Likelihood = -1.97 Transmembrane 78 - 94 (76 - 95)

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.3930 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15 The protein has homology with the following sequences in the databases:
 >GP:CAB15234 GB:Z99120 similar to purine permease [Bacillus subtilis]
 Identities = 216/421 (51%), Positives = 302/421 (71%), Gaps = 5/421 (1%)

20 Query: 6 KQEHSHSQSAVLGLQHVLVSMYAGSILVPIIMIALGYSARELTYLISTDIFMCGVATFLQ 65
 K++H+ Q +LGLQH+L+MYAG+ILVP+++ A+G +A +LTYLI+ D+FMCG AT LQ
 Sbjct: 2 KEQHNAQLQMLLGLQHMLAMYGAILVPLIVGAAIQLNAGQLTYLIAIDLFMCGAATLLQ 61

25 Query: 66 LKLTKHTGVGLPVVLGCAFSQVAPLSIIGAQQGSGAMFGALIASGIYVILVAGIFSFIAR 125
 L ++ G+GLPVVLGC F +V P+ IG+ G A++GA+IA+G+ V+L AG F K+ R
 Sbjct: 62 LWRNRYFGIGLPPVLGCTFTAVGPMISIGSTYGVPAIYGAIIAAGLIVVLAAGFFGKLV 121

30 Query: 126 FFPPIVTGSVITVIGLSLVGVAMGNM--GDNVKE-PTAQSMMLSLTIVIIILLVQKFTKG 182
 FFP+VTG+SV+ +IG+SL+ AM N+ G+ KE + +++L ILL+ F KG
 Sbjct: 122 FFPVVTGSVVMIIGISLIPTAMNLAGGEGSKEFGSLDNVLLGFGVTAFILLLFYFFKG 181

35 Query: 183 FVKISILIGLVAGTLVSAAMGLVDTPVVEASWIHVPTPFYFGMPTFEITSIVMMCI 242
 F++SI+IL+GL+AGT + MG VD + V+EASW+HVP+ FYFG PTFE+ ++V M ++A
 Sbjct: 182 FIRSIAILLGLIAGTAAAYFMGKVDVFSEVLEASWLHVPSLFYFGPPTFELPAVVITMLLVA 241

40 Query: 303 ISGIKTRRPIYYAAGILVIGLLPKFRAMAQMIPSPVLGGAMLVLFMGMVALQGMQLNRV 362
 +S +K+ I ILLV IGL+PK A+ +IP+PVLGGAM+V+FGMV G++ML+ V
 Sbjct: 302 LSKMKSVNVIAITGIILVAIGLVPKAAALITVIPTFVLGGAMIVMFGMVISYGIKMLSSV 361

45 Query: 363 DFQKNEYNFIIAAVSISAGLGFNGT-NLFASLPETAQMFLTNGIIVATLTSVVLNLVLNGK 422
 D ++ N +I A S+S GLG LF+SL A + +GIVI +LT++ L+ K
 Sbjct: 362 DLD-SQGNLLIIASSVSLGLGATTVPALFSSLSGAASVLGSGIVIGSLTAIALHAFQTK 421

An alignment of the GAS and GBS proteins is shown below.

Identities = 328/416 (78%), Positives = 380/416 (90%)

50 Query: 7 SNSQAALLGLQHLLAMYAGSILVPIIMIASALGYNALQTYLIATDIFMCGIATLLQLRLS 66
 S+SQ+A+LGLQH+L+MYAGSILVPIIMIA ALGY+A++LTYLI+TDIFMCG+AT LQL+L+
 Sbjct: 10 SHSQSAVLGLQHVLVSMYAGSILVPIIMIALGYSARELTYLISTDIFMCGVATFLQLKLT 69

55 Query: 67 KHFGVGLFPVVLGCAFSQVAPLSIIGAQQGSGYMFALIASGIYVVLVAGIFSIVANFFPP 126
 KH GVGLFPVVLGCAFSQVAPLSIIGAQQGSG MFGALIASGIYV+LVAGIFS+K+A FPPP
 Sbjct: 70 KHTGVGLFPVVLGCAFSQVAPLSIIGAQQGSGAMFGALIASGIYVILVAGIFSFIARFFPP 129

60 Query: 127 IVTGSVITIGLTLIPVAMGNMGDNKEPSLQSLTSLVTIGVLLINIFAKGFLKSISI 186
 IVTGSVIT IGL+L+ VAMGNMGDN KEP+ QS+ LSL+TI ++LL+ F KGF+KSISI
 Sbjct: 130 IVTGSVITIGLSLVGVAMGNMGDNVKEPTAQSMMLSLTIVIIILLVQKFTKGFVKSISI 189

65 Query: 187 LIGLISGTILAAFMGLVDASVVADAPLVHIPKPFYFGAPRFEFTSILMMCIATVSMVES 246
 LIGL++GT+++A MGLVD + V +A +H+P PFYFG P FE TSI+MMCIATVSMVES
 Sbjct: 190 LIGLVAGTLVSAAMGLVDTPVVEASWIHVPTPFYFGMPTFEITSIVMMCIATVSMVES 249

Query: 247 TGVYLALSDITNDKLDKRLRNGYRSEGLAVLLGGLFNTFPYTGFSONVGLVQISGIRTR 306

-2409-

TGVYLALSD+TND+LD KRLRNGYRSEG+AV LGGLFNTFFPYTGFSQNVGLVQISGI+TR
 Sbjct: 250 TGVYLALSDLTNDQLDEKRLRNGYRSEGI+AVFLGGLFNTFFPYTGFSQNVGLVQISGIKTR 309
 Query: 307 KPIYFTALFLVLGLLPKFGAMAQMIPSPVLGGAMLVLFGMVALQGMKMLNQVDFEHNEH 366
 +PIY+ A LV++GLLPKF AMAQMIPSPVLGGAMLVLFGMVALQGM+MLN+VDF+ NE+
 Sbjct: 310 RPIYYAAGILVVIGLLPKFRAMAQMIPSPVLGGAMLVLFGMVALQGMQMLNRVDFQKNEY 369
 Query: 367 NFIIAAVSIAGVGFGNGTNLFIISLPNTLQMFLTNGIVISTLTAVVLNIIILNGLPKK 422
 NFIIAAVSI+AG+GFNGTNLF SLP T QMFLTNGIVI+TLT+VVLN++LNG K+
 Sbjct: 370 NFIIAAVSIAGLGFNGTNLFIASLPETAQMFLTNGIVIATLTSVVLNLVLNGKDKQ 425

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2140

A DNA sequence (GBSx2256) was identified in *S.agalactiae* <SEQ ID 6611> which encodes the amino acid sequence <SEQ ID 6612>. This protein is predicted to be xanthine phosphoribosyltransferase (xpt). Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1921(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA13587 GB:AJ233894 xanthine phosphoribosyltransferase
 [Streptococcus pneumoniae]
 Identities = 133/162 (82%), Positives = 144/162 (88%)
 Query: 16 GENILKVDSEFLTHQVDFELMQEIGKVFADKYKEAGITKVVVTIEASGIAPAVYAAQALGVP 75
 G+NILKVDSEFLTHQVDF LM+EIGKVFA+K+ AGITKVVVTIEASGIAPA++ A+AL VP
 Sbjct: 1 GDNILKVDSEFLTHQVDFSLMREIGKVFAEKFASAGITKVVVTIEASGIAPALFTAEALNVP 60
 Query: 76 MIFAKKAKNITMTEGILTAEVYSFTKQVTSQVSIVSRFLSNDTIVLIIDDFLANGQAAGK 135
 MIFAKKAKNITM EGILTAEVYSFTKQVTS VSI +FLS +D VLIIDDFLANGQAAGK
 Sbjct: 61 MIFAKKAKNITMTEGILTAEVYSFTKQVTSVSIAGKFLSPEDKVLIIIDDFLANGQAAGK 120
 Query: 136 LLEITGQAGAKVAGIGIVIEKSFQDGRDLLEKTGVPVTSILAR 177
 L++II QAGA V IGIVIEKSFQDGRDLLEK G PV SLAR
 Sbjct: 121 LIQIIEQAGATVEAIGIVIEKSFQDGRDLLEKAGYPVLSILAR 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6613> which encodes the amino acid sequence <SEQ ID 6614>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 156/193 (80%), Positives = 172/193 (88%)
 Query: 1 MKLLEERILKGDVLDGENILKVDSEFLTHQVDFELMQEIGKVFADKYKEAGITKVVVTIEAS 60
 M+LLEERIL DG++LGENILKVD+FLTHQVD+ LM+ IGKVFA KY EAGITKVVVTIEAS
 Sbjct: 1 MQLLEERILTDGNILGENILKVDNFLTHQVDYRLMKAIGKVFAQKYAEAGITKVVVTIEAS 60

-2410-

Query: 61 GIAPAVYAAQALGVPMIFAKKAKNITMTEGILTAEVVSFTKQVTSQVSIVSRFLSND DTV 120
 GIAPAVYAA+A+ VPMIFAKK KNITMTEGILTAEVVSFTKQVTS VSI +FLS +D V
 Sbjct: 61 GIAPAVYAAEAMDVPMIFAKKHKNITMTEGILTAEVVSFTKQVTSVSTAGKFLSKEDKV 120

Query: 121 LIIDDFLANGQAAKGLLEIIGQAGAKVAGIGVIEKSFQDGRDLLEKTGVPVTSLAR IKA 180
 LIIDDFLANGQAAKGL+EIIGQAGA+V G+GIVIEKSFQDGR L+E G+ VTSLAR IKA
 Sbjct: 121 LIIDDFLANGQAAKGLIEIIGQAGAQVVGIVIEKSFQDGRRLIEDMGIEVTSLAR IKN 180

Query: 181 FENGRVVF AEADA 193
 FENG + F EADA
 Sbjct: 181 FENGNLNFLEADA 193

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2141

A DNA sequence (GBSx2257) was identified in *S.agalactiae* <SEQ ID 6615> which encodes the amino acid sequence <SEQ ID 6616>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2546(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15203 GB:Z99120 similar to GMP reductase [Bacillus subtilis]
 Identities = 243/321 (75%), Positives = 286/321 (88%), Gaps = 2/321 (0%)

Query: 7 VFDYEDIQLIPNKCIISRSQADTSVKLGNYTFKLPVIPANMQTIIDEEVAETLACEGYF 66
 VFDYEDIQLIP KCI++SRS+ DTSV+LG +TFKLPV+PANMQTIIDE++A +LA GYF
 Sbjct: 4 VFDYEDIQLIPAKCIVNSRSECDTSVRLGGHTFKLPVVPANMQTIIDEKLAISLAENGYF 63

Query: 67 YIMHRFN EBERKPF IKRMHDKGLIASISVGKDY EYDFVTS LKED--APEFITIDIAHGH 124
 Y+MHRF E R FIK M+ +GL +SISVGVD EY+FY L E+ PE++TIDIAHGH
 Sbjct: 64 YVMHRFEPETRIDFIKDMNARGLFSSISVGKDEEYEFVRQLAEENLTPEYVTIDIAHGH 123

Query: 125 SNSVIEMIQH IKQELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKVK TGF 184
 SN+VIEMIQH+K+ LP++FVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITK+KTGF
 Sbjct: 124 SNAVIEMIQHLLKHLPSFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKIK TGF 183

Query: 185 GTGGWQLAALRWCSKAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHLESPGKL 244
 GTGGWQLAALRWC+KAA KPIIADGGIRTHGDIAKSIRFGA+MVMIGSLFAGH ESPG+
 Sbjct: 184 GTGGWQLAALRWCAKASKPIIADGGIRTHGDIAKSIRFGATMVMIGSLFAGHEESPGQT 243

Query: 245 VEVEGQQPFKEYYGSASEYQKGEHKNVEGK KILLPVKGRLEDTLITEMQODLQSSISYAGGK 304
 +E +G+ +KEY+GSASE+ KGE KNVEGKK+ + KG ++DTL EM+QDLQSSISYAGG
 Sbjct: 244 IEKDGKLYKEYFGSASEFPKGEKKNVEGKKMHVAHKSIKDTLIEMEQDLQSSISYAGGT 303

Query: 305 ELDSL RHVDYVIVKNSIWNGD 325
 +L+++R+VDYVIVKNSI+NGD
 Sbjct: 304 KLNAINRV DYVIVKNSIFNGD 324

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6617> which encodes the amino acid sequence <SEQ ID 6618>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2411-

bacterial cytoplasm --- Certainty=0.2405 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 297/327 (90%), Positives = 311/327 (94%)

Query: 1 MFNDIPVFDYEDIQLIPNKCIISSRSQADTSVKLGNYTFKLPVIPANMQTIIDEEVAETL 60
 MFNDIPVFDYEDIQLIPNKCII+SRQADTSV LG Y FKLPVIPANMQTIIDE +AE L
 10 Sbjct: 8 MFNDIPVFDYEDIQLIPNKCII+SRQADTSVTLGKYQFKLPVIPANMQTIIDETIAEQL 67

Query: 61 ACEGYFYIMHRFNEEERKPFIKRMHDKGLIASISVGKDYDFVTSLKEDAPEFITIDI 120
 A EGYFYIMHRF+E+ RKPFIKRMH++GLIASISVGK EY+FVTSLKEDAPEFITIDI
 15 Sbjct: 68 AKEGYFYIMHRFDEDSRKPFIKRMHEQGLIASISVGKACEYEFVTSLKEDAPEFITIDI 127

Query: 121 AHGHSNSVIEMIQHIKQELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 180
 AHGH+NSVI+MI+HIK ELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV
 20 Sbjct: 128 AHGHANSVIDMIKHIKTELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 187

Query: 181 KTGFGTGGWQLAALRWCSKAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHLES 240
 KTGFGTGGWQLAALRWC+KAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGH ES
 25 Sbjct: 188 KTGFGTGGWQLAALRWCAKAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHFES 247

Query: 241 PGKLVEVEGQQFKEYYGSASEYQKGEHKNVEGKKILLPVKGRLEDTLTEMQQDLQSSISY 300
 PGK VEV+G+ FKEYYGSASEYQKGEHKNVEGKKILLP KG L DTLTEMQQDLQSSISY
 30 Sbjct: 248 PGKTVEVDGETFKEYYGSASEYQKGEHKNVEGKKILLPTKGHLSDTLTEMQQDLQSSISY 307

Query: 301 AGGKELDSLRLHVDYVIVKNSIWNNGDSI 327
 AGGK+LDLRLHVDYVIVKNSIWNNGDSI
 30 Sbjct: 308 AGGKDLDLRLHVDYVIVKNSIWNNGDSI 334

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2142

35 A DNA sequence (GBSx2258) was identified in *S.agalactiae* <SEQ ID 6619> which encodes the amino acid sequence <SEQ ID 6620>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have an uncleavable N-term signal seq

40 INTEGRAL Likelihood = -16.98 Transmembrane 421 - 437 (413 - 443)
 INTEGRAL Likelihood = -8.81 Transmembrane 166 - 182 (159 - 186)
 INTEGRAL Likelihood = -8.55 Transmembrane 220 - 236 (208 - 238)
 INTEGRAL Likelihood = -6.69 Transmembrane 322 - 338 (319 - 353)
 INTEGRAL Likelihood = -5.26 Transmembrane 199 - 215 (196 - 218)
 45 INTEGRAL Likelihood = -4.35 Transmembrane 343 - 359 (342 - 361)
 INTEGRAL Likelihood = -4.09 Transmembrane 291 - 307 (287 - 308)
 INTEGRAL Likelihood = -3.66 Transmembrane 8 - 24 (8 - 27)
 INTEGRAL Likelihood = -3.66 Transmembrane 133 - 149 (133 - 151)
 INTEGRAL Likelihood = -3.19 Transmembrane 254 - 270 (253 - 278)
 50 INTEGRAL Likelihood = -2.50 Transmembrane 53 - 69 (53 - 72)
 INTEGRAL Likelihood = -1.81 Transmembrane 77 - 93 (76 - 95)
 INTEGRAL Likelihood = -1.33 Transmembrane 109 - 125 (109 - 125)

----- Final Results -----

55 bacterial membrane --- Certainty=0.7793 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:CAB61253 GB:AJ250422 ORFC [Oenococcus oeni]
 Identities = 157/447 (35%), Positives = 252/447 (56%), Gaps = 13/447 (2%)

-2412-

Query: 11 AIITTAILGFSGILLETSMNVTFPLLMKEFGVNPVAVIQWVTGNLLAVAVTVPLSAFMK 70
 AI+ A L F G+LIETSMNVTFP LM++F ++ +QW+TT LL VA T+ ++AF+ K
 Sbjct: 15 AILGLAGLAFPCGVLIETSMNVTFPTLMQQFSISLNKVQWLTAYLLLVAAATISIAAFIEK 74

5 Query: 71 NLSEKQIFTLANVLFSLGVLIDSFAPNLAILLVGRVLQGVGTGLALPLLPHIILTQIPME 130
 ++IF A +LF+ GV+ + APN ILL+GR++Q + TGLA+PLL I+ QIP +
 Sbjct: 75 RFIFKKIFFWAGLLFIIGVICSALAPNFIILLIGRLIQALSTGLAIPLLITEIMQQIPQK 134

10 Query: 131 RRGLMMGVAAAMVTLAPAVGPTYGGVISGMLGWKIMIFMLLAPILIISTFIGLASIPKRQV 190
 ++G M + + L P++GPTYGGVI+ L W++IF + PI +I+ IGL+ I ++
 Sbjct: 135 KQGSYMEIVWLLWQPSLGPYGGVITQDLSWRLIFWVFLPIGLIAWLIGLSFIEQKSS 194

15 Query: 191 RINDKLNFPAFISLIGIGLATLLLAIEKMSIF-----YLLVAIVSFVIFYL--NKQ 239
 + FISL + L ++ +A+ I+ +LL+A++ ++F L N +
 Sbjct: 195 PSKIPFAWKQFISLILALLSITVAVNNAGIYGWTSIKFYGFLIAVILLIVFIKLSTNSR 254

20 Query: 240 LEFLNLNVFKDKDFSILLYGVLAFOMPIPLALSFLLPNLLQLVLHQTSTKAGLFMFPGAIA 299
 +++++FK +F L Q I L+L+FLLPN QL+L + +G+ + G++
 Sbjct: 255 QALISISIFKKWEFVCPLLIYFLIQFIQLSLTFLLPNYAQLILKKGVMISGIMLLCGSLI 314

25 Query: 300 VVFLSPFAGYLLDKIGAFKPIIMIGISLSLIGLIGTAIFIPAKSVVLLAFDILTKIGMGI 359
 L P G +LD P++IG + I IF SV ++ A ++ IG
 Sbjct: 315 SAILQPLTGRMLDSFSVKIPLVIGAFFLITSTISFTIFQRYLSVFLIAALYVIYMIGFSF 374

30 Query: 360 GASNMVTTALTCLKPAQSADGNSILNTLQQFAGAFATAVASQIFTIGQVAIPKNGAIIGS 419
 +N +T AL KL +DGN++ NTLQQ+AG+ T+VAS + G K GS
 Sbjct: 375 VFNNSLTYALQKLEPLKLISDGNVFNLTQQYAGSLGTSVASALLANGIGTDGKQSNYTG 434

Query: 420 Q--FAVLFVIVVILAIIVGLTYLRKRK 444
 + F + F+ +++ ++ +K K
 Sbjct: 435 RHIFILNFISCAIVVILIFSIQRKKNK 461

There is also homology to SEQ ID 46.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2143

A DNA sequence (GBSx2259) was identified in *S.galactiae* <SEQ ID 6621> which encodes the amino acid sequence <SEQ ID 6622>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2151(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6595> which encodes the amino acid sequence <SEQ ID 6596>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2413-

Identities = 74/214 (34%), Positives = 112/214 (51%), Gaps = 5/214 (2%)

Query: 13 NESENFFITLKTYFNLFISIQTIT--DISTLNHADFDGSAFHDIETSIPLVIDSNY 69
 N+ E F L +F++LF + I+T +I + + F G F+FH+ + +P L ++
 5 Sbjct: 15 NQLEETFIRESLHHFSLFEVTILTISKANIQSNQLSTFQGIFSFHEHDIDLPTLYFKTSQ 74

Query: 70 LAISQTSNKIEANDIKTFSELSKTMTEFHYMLNFDLNFHLPYRFRHLNKGQTIYSNHKP 129
 ++ + LS+ +T F+ + +LP + RL + +G I NH
 10 Sbjct: 75 HGQGFLVTESVFDQATAVLSLSQYLTFQKFDGHFLQYLPQARLSDANGNIIVDNHAF 134

Query: 130 EDPFDIYPEEEYPIDKWVQNSLIEKKAKELHLLLPASQDYILVQSYKRLENDSGQLVG 189
 F P + I+ W+ L LLPS S D+I +Q Y+ L+N GQLVG
 15 Sbjct: 135 NGSF--LPTTDKETEDWILABRLSDNPCKTFLPLPSGSLDHIYMQHYQALKNPQGQLVG 192

Query: 190 IEHVHNKIPLLLEGYLKESGQAIVGWSVDVTSGASI 223
 ++ V +IKPLL YL+E+GQAIVGWSVDVTSG SI
 15 Sbjct: 193 LDTVQDIKPLLNQYLEETGQAIVGWSVDVTSGPSI 226

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 20 vaccines or diagnostics.

Example 2144

A DNA sequence (GBSx2260) was identified in *S.galactiae* <SEQ ID 6623> which encodes the amino acid sequence <SEQ ID 6624>. Analysis of this protein sequence reveals the following:

Possible site: 18
 25 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.10	Transmembrane	431 - 447 (423 - 452)
INTEGRAL	Likelihood = -8.92	Transmembrane	149 - 165 (147 - 174)
INTEGRAL	Likelihood = -8.86	Transmembrane	404 - 420 (402 - 428)
INTEGRAL	Likelihood = -7.91	Transmembrane	299 - 315 (293 - 318)
30 INTEGRAL	Likelihood = -6.42	Transmembrane	380 - 396 (374 - 398)
INTEGRAL	Likelihood = -5.31	Transmembrane	350 - 366 (347 - 367)
INTEGRAL	Likelihood = -4.57	Transmembrane	56 - 72 (54 - 74)
INTEGRAL	Likelihood = -3.24	Transmembrane	172 - 188 (171 - 198)
INTEGRAL	Likelihood = -1.33	Transmembrane	224 - 240 (224 - 240)
35 INTEGRAL	Likelihood = -0.59	Transmembrane	101 - 117 (101 - 117)

----- Final Results -----
 bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF84709 GB:AE004010 potassium uptake protein [Xylella
 fastidiosa]
 45 Identities = 201/570 (35%), Positives = 319/570 (55%), Gaps = 34/570 (5%)

Query: 1 MAEMQHVNHSSFDKASKAGFII--ALGIVYGDIGTSPLYTMQSLVENQGGISSVTESFIL 58
 M+ H + ++ G II A+G+V+GDIGTSPLYT++ G++ ++ +L
 50 Sbjct: 1 MSTSSHSGDCTAVPSNSNGTIIILSAIGVVFEGDIGTSPLYTLKEAFSPNYGLTPNHDT-VL 59

Query: 59 GSISLIWTLTLITTIKYVLVALKADNHHHEGGIFSLYTLVRKMTFW-----LIVPAVI 111
 G +SLI W + L+ TIKYV V ++ DN EGGI +L L ++ P+ + + +
 55 Sbjct: 60 GILSLIFWAMMLVVITIKYVAVIMRVDNDGEGGIMALTALTQRTMPFGSRSIYIVGILGIF 119

Query: 112 GGATLLSDGALTPAVTVTSAVEGLKVVPVSLQHIFQNSNVIFATLFIILLFFAIQRFGTG 171
 G + DG +TPA++V SAVEGL+V F V+ TL +L+LLF QRFGT
 60 Sbjct: 120 GTSLFFGDGVITPAISVLSAVEGLEVAEPHMKAF-----VVPITLAVLILLFLCQRFGT 174

Query: 172 VIGKLFPGPIMFIWFAFLGISGLLNSFAHPEVFKAINPYGLKLLFSPENHKGIFILGSIF 231
 +GK FGPI +WF +G+ G+ N PEV AINP +GL F +F+LG++
 60 Sbjct: 175 RVGKTFGPITLLWFIAIGVGVYINIAQAPEVLHAINPSWGLH-FFLEHGWHSMPVLGAVV 233

-2414-

Query: 232 LATTGAEALYSDLGHVGRGNIHVSWPFVKVAII-LSYCGQGAWILANKNAGNELNPFASF 290
 LA TG EALY+D+GH G I +W +V + ++ L+Y GQGA +L+N A NPF+ S
 Sbjct: 234 LAVTGGEALYADMGHFGAKAIRHAWMYVVLPLMLALNYLGQGAALVLSNPTAIG--NPFYQS 291

5 Query: 291 IPSQFTMHVVILATLAAIIASQALISGSFTLVSEAMRLKIFPQFRSTYPGDN-IGQTYIP 349
 IP ++ LAT AA+IASQALI+GS++L S+AM+L P+ + + IGQ Y+P
 Sbjct: 292 IPDWGLYPMIALATAAAVIASQALITGSYSLSQAMQLGYIPRMNVRHTSQSTIGQIYVP 351

10 Query: 350 VINWFLFAITTSIVLLFKTSAHMEAYGLAITITMLMTTILLSFFL-IQKGVKRGVLVLLM 408
 +NW L + V+ F S M +AYG+A+T TM++TT+L+ + V R ++ +M
 Sbjct: 352 TVNWTLTTLVILTVIGFGDSTSMASAYGVAVTGTMMITTVLMIYARANPRVPRMLWMM 411

15 Query: 409 MIFFGILEGIFFLASAVKFMHGGYVVVIIAVAIIFIMTIWYKGSKIVSRVYKL--LDLKD 466
 I F ++G FF A+ +KFM G + +++ V I M W +G K++ ++ ++L +
 Sbjct: 412 AIVFIAVDGAFFYANIIKFMGDGAWFPLLLGVVIFTFMRTWLRGRKLLHEEMRKDGINLDN 471

20 Query: 467 YIGQLDKLRHDHRYPIYHTNVVYLTNRMEEDMIDKSIMYSILDKRPKKAQVYWFVNIKVT 526
 ++ L L + P V+LT + ++ ++M+++ + + F+ +K
 Sbjct: 472 FLPGGL-MLAPPVKVP---GTAFLIT--ADSTVVPHALMHNKHNKVLHERNV-FLTVKTL 524

Query: 527 DEPYTA---EYKVDMMGTDFIVKVELYLGF 553
 PY A K++ + F +V + GF
 Sbjct: 525 KIPYAANSERLKIEPISNGF-YRVHIRFGF 553

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6625> which encodes the amino acid sequence <SEQ ID 6626>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq

30	INTEGRAL	Likelihood = -11.78	Transmembrane	428 - 444 (421 - 453)
	INTEGRAL	Likelihood = -8.70	Transmembrane	146 - 162 (144 - 171)
	INTEGRAL	Likelihood = -7.64	Transmembrane	404 - 420 (398 - 426)
	INTEGRAL	Likelihood = -4.88	Transmembrane	296 - 312 (294 - 315)
	INTEGRAL	Likelihood = -4.57	Transmembrane	53 - 69 (51 - 71)
	INTEGRAL	Likelihood = -3.93	Transmembrane	347 - 363 (343 - 363)
35	INTEGRAL	Likelihood = -2.50	Transmembrane	372 - 388 (371 - 388)
	INTEGRAL	Likelihood = -1.33	Transmembrane	169 - 185 (169 - 185)
	INTEGRAL	Likelihood = -1.33	Transmembrane	221 - 237 (221 - 237)

----- Final Results -----

40	bacterial membrane	--- Certainty=0.5713 (Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000 (Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:AAF84709 GB:AE004010 potassium uptake protein [Xylella
 fastidiosal]
 Identities = 177/467 (37%), Positives = 270/467 (56%), Gaps = 20/467 (4%)

50 Query: 7 TAFDKASKAGFII-ALGIVYGDIGTSPLYTIQSLVENQGGVNQVSESFILGSISLIWTL 65
 TA S I+ A+G+V+GDIGTSPLYT++ G+ ++ +LG +SLI W +
 Sbjct: 11 TAVPSNSNGTIILSAIGVVFGDIGTSPLYTLKEAFSPNYGLTPNHDT-VLGILSLIFWAM 69

55 Query: 66 TLITTIKYVLIALKADNHHGGIFSLFTLVKRMSPW-----LIIPAMIGCATLLSDGA 118
 L+ TIKYV + ++ DN EGGI +L L ++ P+ + I + G + DG
 Sbjct: 70 MLVVITIKYVAVIMRVDNDGEGGIMALTALTQRTMPFGSRSIYIVGILGIFGTSLFPFGDG 129

60 Query: 119 LTPAVTVTSAIEGLKAVPGLSHIYQNQTINVIITLVILIVLFGIQRFGTGFQKIFGPVM 178
 +TPA++V SA+EGL+ + V+ TL +LI+LF QRFGT +GK FGP+
 Sbjct: 130 ITPAISVLSAVEGLEVAEPHMKAF-----VVPITLAVLILLFLCQRFQTERVKGTFGPIT 184

65 Query: 179 FIWFSFLGVSGFFNTLGHLEIFKAINPYALHLLFSPENHRGIFILGSIFLATTGAEALY 238
 +WF +GV G +N E+ AINP + LH F +F+LG++ LA TG EALY
 Sbjct: 185 LLWFITAIGVGVYNIAQAPEVLHAINPSWGLH-FFLEHGWHSMFVLGAUVLAVTGGEALY 243

Query: 239 SDLGHVGRGNIYVSWPFVKM-CIVLSYCGQAAILANKHSGIELNPFASFVPSQLRVYLV 297

-2415-

+D+GH G I +W +V + + L+Y GQ A +L+N + NPF+ S+P ++
 Sbjct: 244 ADMGHFGAKAIRHAWMYVVLMLALNLYLQGQALVLSNFTA--IGNPFYQSIPDWGLYPMI 301

Query: 298 SLATLAAIIASQALISGSFTLVSEAMRLKIFPLFRVTYPG-ANLGQLYIPVINWILFAVT 356
 +LAT AA+IASQALI+GS++L S+AM+L P V + + +GQ+Y+P +NW L +
 Sbjct: 302 ALATAAAVIASQALITGSYSLSQAMQLGYIPRNMVRHTSQSTIGQIYVPTVNWTLTLTV 361

Query: 357 SCTVLAFTSAHMEAAAYGLAITITMLMTTILLKYLIKKGTRPILAHLVMAF-FALVEFI 415
 TV+ F S M +AYG+A+T TM++TT+L+ Y P L +MA F V+
 Sbjct: 362 ILTVIGFGDSTSMASAYGVAVTGTMITTVMIIYARANPRVPRMLMWMMAIVFIAVDGA 421

Query: 416 FFLASAIKFMHGGYAVVILALAIVFVFMFIWHAGTRIVFKYVKSINLN 462
 FF A+ IKFM G + ++L + I M W G +++ + ++ +N
 Sbjct: 422 FFYANIIKFMGAWFPLLLGVVIFTFMRTWLGRKLLHEEMRKDGIN 468

An alignment of the GAS and GBS proteins is shown below.

Identities = 485/651 (74%), Positives = 575/651 (87%)

Query: 10 SSFDKASKAGFIIALGIVYGDIGTSPLYTMQSLVENQGGISSVTESFILGSISLIWTLT 69
 ++FDKASKAGFIIALGIVYGDIGTSPLYT+QSLVENQGG++ V+ESFILGSISLIWTLT
 Sbjct: 7 TAFDKASKAGFIIALGIVYGDIGTSPLYTIQSLVENQGGVNVQSESFILGSISLIWTLT 66

Query: 70 LITTIKYVLVALKADNHHEGGIFSLYTLVRKMTPLIVPAVIGGATLLSDGALTPAVTVT 129
 LITTIKYVL+ALKADNHHEGGIFSL+TLVRKM+PWLI+PA+IGGATLLSDGALTPAVTVT
 Sbjct: 67 LITTIKYVLIALKADNHHEGGIFSLFTLVRKMSFWLIIPAMIGGATLLSDGALTPAVTVT 126

Query: 130 SAVEGLKVPSLQHFQNSNVIFATLFIILLFLLFAIQRFGTGVIGKLFPGPIMFIWFAFLG 189
 SA+EGLK VP L HI+QNQ+NVI TL IL++LF IQRFGTG IGK+FGP+MFIWF+FLG
 Sbjct: 127 SAIEGLKAVPGLSHIYQNQNTVITTLVILIVLFGIQRFGTGFIGKIFGPVPMFIWFSFLG 186

Query: 190 ISGLINSFAHPEVFKAINPYGYGLKLLFSPENHKGIFILGSIFLATTGAEALYSDLGHVGR 249
 +SG N+ H E+FKAINPY L LLFSPENH+GIFILGSIFLATTGAEALYSDLGHVGR
 Sbjct: 187 VSGFFNTLGHLEIFKAINPYALHLLFSPENHRGIFILGSIFLATTGAEALYSDLGHVGR 246

Query: 250 GNIHVSWPVFKVAIILSYCGQAWILANKNAGNELNPPFASIPSQFTMHVILATLAAII 309
 GNI+VSWPFVK+ I+LSYCGQ AWILANK++G ELNPPFAS+PSQ +++V LATLAAII
 Sbjct: 247 GNIYVSWPFVKMCIVLSYCGQAAILANKHSGIELNPPFASVPSQLRVYLVSLATLAAII 306

Query: 310 ASQALISGSFTLVSEAMRLKIFPQFRSTYPGDNIGQTYIPVINWFLFAITTSIVLLFKTS 369
 ASQALISGSFTLVSEAMRLKIFP FR TYPG N+GQ YIPVINW LFA+T+ VL F+TS
 Sbjct: 307 ASQALISGSFTLVSEAMRLKIFPLFRVTYPGANLGQLYIPVINWILFAVTSCTVLAFTS 366

Query: 370 AHMEAAAYGLAITITMLMTTILLSFPLIQGVKRGVLVLLMIFFGILEGIFFLASAVKFMH 429
 AHMEAAAYGLAITITMLMTTILL ++LI+KG + L L+M FF ++E IFFLASA+KFMH
 Sbjct: 367 AHMEAAAYGLAITITMLMTTILLKYLIKKGTRPILAHLVMAFFALVEFIFFLASAIKFMH 426

Query: 430 GGYVVVIIAVAIIFIMTIWYKGSKIYVSRVKKLLDKDYIGQLDKLRHHRYPYHTNVVY 489
 GGY VVI+A+AI+F+M IW+ G++IV +YVK L+L DY Q+ +LR D + +Y TNVVY
 Sbjct: 427 GGYAVVILALAIVFVFMFIWHAGTRIVFKYVKSINLNDYKEQIKQLRDDVCFDLYQTNVVY 486

Query: 490 LTNRMEEDMIDKSIMYSILDKRPPKAQVYWFVNIVKVTDEPYTAEYKVDMMGTDVIVKVEL 549
 L+NRM++ MID+SI+YSILDKRPP+AQVYWFVN++VTDEPYTA+YKVDMMGTD++V+V L
 Sbjct: 487 LSNRMQDHMIDRSILYSILDKRPPKAQVYWFVNIVKVTDEPYTAKYKVDMMGTDYVVRVNL 546

Query: 550 YLGFKMRQTVSRYLRTIVEELLESGRLPKQKTYSVRFDNSVGDVFRFIVLDERFSSSQNL 609
 YLGF+M QTV RYLRTIV++L+ESGRLPKQ + Y++ P +VGDFRF++++ER S+++ L
 Sbjct: 547 YLGRMPQTVPRYLRTIVQDLMESGRLPKQEYETITPGRDVGDFRFVLEERVSNARQL 606

Query: 610 KPGERFVMLMKSSIKHWTATPIRWFLQFSEVTTEVPLIFTANRGLPIKE 660
 ERF+M K+SIKH TA+P+RWFLQ+SEVT EVVPLI + LPIKE
 Sbjct: 607 SNFERFIMQTKASIKHVTASPMRWFLQYSEVTLEVPLILSDVLKLPKE 657

A related GBS gene <SEQ ID 8983> and protein <SEQ ID 8984> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: 5.84
GvH: Signal Score (-7.5): -4.59
Possible site: 18

5	ALOM program	count: 10	value: -12.10	threshold: 0.0
	INTEGRAL	Likelihood = -12.10	Transmembrane	431 - 447 (423 - 452)
	INTEGRAL	Likelihood = -8.92	Transmembrane	149 - 165 (147 - 174)
	INTEGRAL	Likelihood = -8.86	Transmembrane	404 - 420 (402 - 428)
	INTEGRAL	Likelihood = -7.91	Transmembrane	299 - 315 (293 - 318)
10	INTEGRAL	Likelihood = -6.42	Transmembrane	380 - 396 (374 - 398)
	INTEGRAL	Likelihood = -5.31	Transmembrane	350 - 366 (347 - 367)
	INTEGRAL	Likelihood = -4.57	Transmembrane	56 - 72 (54 - 74)
	INTEGRAL	Likelihood = -3.24	Transmembrane	172 - 188 (171 - 198)
	INTEGRAL	Likelihood = -1.33	Transmembrane	224 - 240 (224 - 240)
15	INTEGRAL	Likelihood = -0.59	Transmembrane	101 - 117 (101 - 117)
	PERIPHERAL	Likelihood = 0.85	20	
	modified ALOM score: 2.92			

```

----- Reasoning Step: 3
20
----- Final Results -----

```

25

ORF02578(367 - 1680 of 2607)
 GP|9106998|gb|AAF84709.1|AE004010_6|AE004010(25 - 463 of 634) potassium uptake protein
 {Xylella fastidiosa}
 %Match = 17.8
 %Identity = 40.4 %Similarity = 63.7
 Matches = 177 Mismatches = 150 Conservative Sub.s = 102

40 420 450 480 510 540 570 600 612
IGTSPLYTMQSLVENQGGISSVTESFILGSISLIWTLTLITTIKYVLVALKADNHHEGGIFSLYTLVRKMTP-----W
|||||::: :: :|| :|| | : |||| : : || ||| :| | : : |
IGTSPLYTLKEAFSPNYGLTPNHD-TVLGILSLIFWAMMLVVTIKYVAVIMRVDNDGEGGIMALTALTORTMPFGSRSIY
 50 70 80 90 100 110

879 909 939 969 999 1029 1059 1089
WFAFLGISGLLNSFAHPEVFKAINPYYGKLLFSPENHKGIFILGSIFLATTTGAELYSIDLGHVGRGNIHVSWPFKVKVAI
|| :|: | : |||: ||| :|| :: | :|:||: | | ||| |:| :| | :| :| :| :
55 WFTAIGVVGVYNTAQAEVLHAINPWSWGLHFLEHGWS-MFVLGAVVLAVTGGEALYADMGHFGAKAIRHAMWYVVLPM

200 210 220 230 240 250 260

65 1353 1383 1413 1443 1473 1500 1530 1560
-IGQTYIPVINWFLFAITTSIVLLFKTSAHMEAAAYGLAITTTMLMTTILLSFFD-IQKGVKRGVLVLLMMIFFGILEGIF
||| :| :|| : : | : | | :||| :| || :||| : : | : : :| | | :|| ||
TIGQIYVPTVNNWTLTLVILTVIGEGDSTSMASAYGVAVTGMMITTTVLMIIYARANPRVPRMLMMMAIVETAVDGAFF

-2417-

360 370 380 390 400 410 420

1590 1620 1650 1680 1710 1740 1770 1800

5 LASAVKFMHGGYVVVIIAVAIIFIMTIWYKGSKIVSRVVKLLDLKDYIGQLDKLRHDHRYPIYHTNVVYLTNRMEEDMID

| : : || | : : : | | | : | : : : :

YANIIKFMDGAWFPLLGLGVVITFMRTWLRGRKLLHEEMRKDGINLDNFLPGLMLAPPVKVPGTAVFLTADSTVVFHALM

440 450 460 470 480 490 500

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2145

A DNA sequence (GBSx2261) was identified in *S.agalactiae* <SEQ ID 6627> which encodes the amino acid sequence <SEQ ID 6628>. This protein is predicted to be serine dehydrogenase. Analysis of this protein sequence reveals the following:

15 Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3261(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAD07424 GB:AE000552 short chain alcohol dehydrogenase
[Helicobacter pylori 26695]
Identities = 18/31 (58%), Positives = 25/31 (80%)

Query: 3 WVASQPEHININRIEIMPVSQTYGQPQPVYRD 33
W+ QP H+NINRIEIMP+SQT+ P P +++

30 Sbjct: 219 WIYEQPLHVNINRIEIMPISQTFAPLPTHKN 249

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6629> which encodes the amino acid sequence <SEQ ID 6630>. Analysis of this protein sequence reveals the following:

35 Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1021(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 24/33 (72%), Positives = 29/33 (87%)

45 Query: 1 MSWVASQPEHININRIEIMPVSQTYGQPQPVYRD 33
+SWV QP H+N+NRIE+MPVSQ+YGPQPV RD

Sbjct: 20 VSWVIHQPPHVNINRIEIMPVSQSYGQPQVTRD 52

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2146

A DNA sequence (GBSx2262) was identified in *S.agalactiae* <SEQ ID 6631> which encodes the amino acid sequence <SEQ ID 6632>. Analysis of this protein sequence reveals the following:

-2418-

Possible site: 21
>>> May be a lipoprotein

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9337> which encodes amino acid sequence <SEQ ID 9338>
10 was also identified. A further related GBS nucleic acid sequence <SEQ ID 10781> which encodes amino
acid sequence <SEQ ID 10782> was also identified. A further related GBS nucleic acid sequence <SEQ ID
10951> which encodes amino acid sequence <SEQ ID 10952> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAA32349 GB:X14130 ORF (AA 1 to 299) [Lactococcus lactis subsp.
 cremoris]
 Identities = 72/215 (33%), Positives = 110/215 (50%), Gaps = 8/215 (3%)

Query: 4 RSKLAAGFLTLMSVATLAACSGKTSNGTN--VVTMKGDTITVSDFYDQVKTSKAAQQSML 61
 + K+ L + L SG SN T+ V T G +T S FY ++K S + +
20 Sbjet: 2 KKKMLRKVLLASTATALLLLSGCQSNQTDQTVATYSGGKVTESSFYKELKQSPTTKTMLA 61

Query: 62 TLILSRVFDTPQYGDVKVSEAYNKTAKGYGNSFSSALSQAGLTPEGYKQQIRTTMLV 121
 +++ R + YG VS K V++AY+ + YG +F + LSQ G + +K+ +RT L
25 Sbjet: 62 NMLIYRALNHAYGKSVSTKTVDAYDSYKQYGENFDAFLSQNGFSRSSFKESLRTNFLS 121

Query: 122 EYAVKEAAKKELTEANYKEAYKNYTPETSVQVIKLDKADGADFAKIAK 181
 E A+K+ K+++E+ K A+K Y P+ +VQ I ED AK V+ D+ A G DFA +AK
Sbjet: 122 EVALKKL--KKVSESQLKAANKTYQPKVTQVHILTSDEDTAKQVISDLAA-GKDFAMLA 178

30 Query: 182 E---KTTATDKKVEYKFDSAGTTLPEKVMASAAFKL 213
 T D + F+ TL AA+KL
Sbjet: 179 TDSIDTATKDNKGKISFELNNKTLDATFKDAAYKL 213

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6633> which encodes the amino acid
35 sequence <SEQ ID 6634>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> May be a lipoprotein

40 ----- Final Results -----

 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the databases:

>GP:AAA25247 GB:M83946 maturation protein [Lactobacillus paracasei]
 Identities = 88/294 (29%), Positives = 146/294 (48%), Gaps = 14/294 (4%)

Query: 7 LIASVVTLASVMAAACQSTNDNTKVISMKGDTISVSDFYNETKNTEVSQKAMLNLVISR 66
 L+AS T +++ L+ CQS + KV + G ++ S+FY E K + ++ + N++I R
50 Sbjet: 10 LLASTAT--ALLLLSGCQSNQADQKVATYSGGKVTESNFYKELKQSPTTKTMLANMLIYR 67

Query: 67 VFEAQYGDVKVSKKEVEKAYHKTAEQYGASFSAAALQSSLTPETFKRQIRSSKLVEYAVKE 126
 YG VS K V AY +QYG +F A L+Q+ + +FK +R++ L E A+K+
55 Sbjet: 68 ALNHAYGKSVSTKTVDAYDSYKQYGENFDAFLSQNGFSRSSFKESLRTNFLSEVALKK 127

Query: 127 AAKKELTTQEQYKAYESYTPMAVEMITLDNEETAKSVLEELKARGADFTAIKE---KT 183
 K+++ + K +++Y P + V+ I +E+TAK V+ +L A G DF +AK T
60 Sbjet: 128 L--KKVSESQLKAVWKTYQPKVTQVHILTSDEDTAKQVISDL--AAGKDFATLAKTDSIDT 184

Query: 184 TTPEKKVITYKFDGATNVPTDVVKAASSLNEGGISDVISVLDPTSYQKKFYIVKVTKKAE 243

-2419-

T + F+S + AA L G + P + ++K+
 .Sbjct: 185 ATKDNGGKISFESNNKTLDATFKDAAYKLKNGDYTQT-----PVKVTNGYEVIKMINH-P 238
 Query: 244 KKS DWQ EYKRLKAI IIAE KSKD M N FQNKV IANALDKANVKIKDKAFANILAQY 297
 5 K + KK L A + A+ S+D + +VI+ L +V IKDK A+ L Y
 Sbjct: 239 AKGTF TSSKKALTASVYAKWSRDSSIMQ RVISQVLKNQHVTIKDKLADALDSY 292

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/213 (58%), Positives = 168/213 (78%), Gaps = 1/213 (0%)
 10 Query: 1 MKTRSKLAAGFLTLMSVATLAACSGKTSNGTNVVTMKGDTITVSDFYDQVKTSKAAQQSM 60
 MK +KL A +TL SV LAAC T++ T V++MKGDTI+VSDFY++ K ++ +Q++M
 Sbjct: 1 MKNSNKL IASVVTLASVMALAACQS-TNDNTKVISMKGDTISVSDFYNETKNTEVSQKAM 59
 15 Query: 61 LTLILSRVFD TQYGD KVS DKKVSEAYNKTAKGYGNSFSSALSQAGLTPEGYKQQIRTTML 120
 L L++SRVF+ QYGD KVS K+V +AY+KTA+ YG SFS+AL+Q+ LTPE +K+QIR++ L
 Sbjct: 60 LNLVISRVFEAQYGD KVS KEVEKAYHKTAEQYGASFSAAALQSSLTPE TFKRQIRSSKL 119
 20 Query: 121 VEYAVKEAAKKELTEANYKEAYKNYTPETSVQVIK LDAEDKAKSVLKDV KADGADF AKIA 180
 VEYAVKEAAKKELT YK+AY++YTP +V++I LD E+ AKSVL+++KA+GADF IA
 Sbjct: 120 VEYAVKEAAKKELTTQ EYKAYESYTP TMAVEMITLDNEETAKSVLEELKAE GADFT AIA 179
 Query: 181 KEKTTATDKKVEYKFDSAGTTL PKEVMSAAFKL 213
 KEKTT +KKV YKFDS T +P +V+ AA L
 25 Sbjct: 180 KEKTTTPEKKV TYKFDSGATNVPTD VVKAASSL 212

SEQ ID 10782 (GBS657) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 8-10; MW 62.8kDa) and in Figure 187 (lane 3; MW 63kDa). Purified GBS657-GST is shown in Figure 245, lanes 2 & 3.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2147

A DNA sequence (GBSx2263) was identified in *Sagalactiae* <SEQ ID 6635> which encodes the amino acid sequence <SEQ ID 6636>. This protein is predicted to be methyltransferase. Analysis of this protein
 35 sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]
 Identities = 132/227 (58%), Positives = 169/227 (74%)
 Query: 1 MVQSYSKNANHNMRPVPVKEEIVQYMRQH QKQNNGLAELEAF AKQENIPIPHETATYF 60
 MV++Y +N M RPVVK E+V++MR Q Q G LAE+ FAK+ NIP+IPHET YF
 50 Sbjct: 1 MVETYKSTSNPMMNRPVVKAELVEWMRSSQTQVIGELAEVLNFAKENNIPVIPHETVLYF 60
 Query: 61 RFLMQTLQPKHILEIGTAIGFSALLMAENAPEAKIT TIDRNEEMIALAKENFAKYDHNQ 120
 + L+ L+PK ILEIGTAIGFSAL+MA+ PEA+I TIDRN EMI LAK+N AKYD+ NQ
 Sbjct: 61 QMLLSLLKPKRILEIGTAIGFSALVMAQEVPEAEIVTIDRNP EMIELAKKNLAKYDHRNQ 120
 55 Query: 121 ITLLEGDAVDVLQTL DKS YDFVFMDSAKSKYIVFLPQVLKHL DVGGVVVLDDIFQGGDIA 180
 I L EGDA DVLQ L +D VFMSAKSKY+ FLP+ L+ L G+++DD+ FQ G+I
 Sbjct: 121 IQLKEGDAADVLQELKGPFDLVFMDSAKSKYVEFLPKSLELLSENGLILMDDVFQAGEIL 180

-2420-

Query: 181 KPIDEVRRGQRTTIYRGLQRLFDSTLQHPDLTATLVPLGDGLLMIRKN 227
 PI EV+R QR + RGL++LFD +P +++PLGDGLLMI+K+
 Sbjet: 181 LPIMEVKRNQRALERGLRKLDFDEVFDNPKYMTSVLPLGDGLLMIKKH 227

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6637> which encodes the amino acid sequence <SEQ ID 6638>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 153 - 169 (152 - 170)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]
 Identities = 134/227 (59%), Positives = 169/227 (74%)
 Query: 1 MVKSYSKTANHNMRPVVKEELVHYMRTRQKQTTFGLAELEQFARQENIPIIQPEVVAYF 60
 MV++Y T+N M RPVVK ELV +MR+ Q Q TG LAE+ FA++ NIP+I E V YF
 Sbjet: 1 MVETYKSTSNPMMNRPVVKAELVEWMRSSQTQVTGELAEVLNFAKENNIPVIPHETVLYF 60
 Query: 61 RFLQLSQPKHILEIGTAIGFSALLMAENAPDATIVTIDRNREMIIDFAKANFAKYDSRQQ 120
 + LL L+PK ILEIGTAIGFSAL+MA+ P+A IVTIDRN EMI+ AK N AKYD R Q
 Sbjet: 61 QMLLSLLKPKRILEIGTAIGFSALVMAQEVPEAEIVTIDRNPEMIELAKNLAKYDHRNQ 120
 Query: 121 IRLLEGDAADILSTLEGNFDFVFMDSAKSKYIVFLPEILRLKVGGVVILDDVFQGGDIT 180
 I+L EGDAAD+L L+G FD VFMDSAKSKY+ FLP+ L LL G+++DDVFQ G+I
 Sbjet: 121 IQLKEGDAADVQLQELKGPFDLVFMDSAKSKYVEFLPKSLELLSENGLILMDDVFQAGEIL 180
 Query: 181 KPIEDIRRGQRTTIYRGLQSLFDATLTHPNLTTSVLPLSDGLLMIRKN 227
 PI +++R QR + RGL+ LFD +P TS++PL DGLLMI+K+
 Sbjet: 181 LPIMEVKRNQRALERGLRKLDFDEVFDNPKYMTSVLPLGDGLLMIKKH 227

An alignment of the GAS and GBS proteins is shown below.

Identities = 177/235 (75%), Positives = 199/235 (84%)
 Query: 1 MVQSYSKNANHNMRPVVKEEIVQYMRQHQQNNGCLAELEAFQENIPIIPHETATYF 60
 MV+SYSK ANHNMRPVVKEE+V YMR QKQ G LAELE FA+QENIPII E YF
 Sbjet: 1 MVKSYSKTANHNMRPVVKEELVHYMRTRQKQTTFGLAELEQFARQENIPIIQPEVVAYF 60
 Query: 61 RFLMQTLQPKHILEIGTAIGFSALLMAENAPEAKITTIDRNEEMIALAKENFAKYDNHNQ 120
 RFL+Q+LQPKHILEIGTAIGFSALLMAENAP+A I TIDRN EMI AK NFAKYD+ Q
 Sbjet: 61 RFLQLSQPKHILEIGTAIGFSALLMAENAPDATIVTIDRNREMIIDFAKANFAKYDSRQQ 120
 Query: 121 IRLLEGDAVDVLQTLKSYDFVFMDSAKSKYIVFLPQVLKHLVDVGGVVVLLDDIFQGGDIA 180
 I LLEGDA D+L TL+ ++DFVFMDSAKSKYIVFLP++L+ L VGGVV+LDD+FGGGDI
 Sbjet: 121 IRLLEGDAADILSTLEGNFDFVFMDSAKSKYIVFLPEILRLKVGGVVILDDVFQGGDIT 180
 Query: 181 KPIDEVRRGQRTTIYRGLQRLFDSTLQHPDLTATLVPLGDGLLMIRKNADHIVLED 235
 KPI+++RRGQRTTIYRGLQ LFD+TL HP+LT +LVPL DGLLMIRKN IVL D
 Sbjet: 181 KPIEDIRRGQRTTIYRGLQSLFDATLTHPNLTTSVLPLSDGLLMIRKNQADIVLDP 235

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2421-

Example 2148

A DNA sequence (GBSx2264) was identified in *S.agalactiae* <SEQ ID 6639> which encodes the amino acid sequence <SEQ ID 6640>. This protein is predicted to be phosphoglycolate phosphatase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 50
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2193 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8985> which encodes amino acid sequence <SEQ ID 8986> was also identified. This protein appears to be a hydrolase *i.e.* an exposed protein.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]
  Identities = 39/117 (33%), Positives = 67/117 (56%), Gaps = 9/117 (7%)

Query:  98  KEQESRDSKIHLM-PYAKEILEWTKQDIPNPMYTHKGASTHSVLETLQISHYFDEILTG 156
20      KE E+R+ +  ++      ++LE  Q  +F+ +H+      +LE  I+ YF E++T
Sbjct:  25  KENEARELEHPILFEGVSDLLEDILNQGGRHFLVSHRNDQVLEILEKTSIAAYFTEVVTS 84

Query:  157  VSGFERKPHQPQGINYLVKRYSLDKSMYYIGDRPLDLEVAQNAGIKS-----INLR 207
25      SGF+RKP+P+ + YL ++Y +  +  IGD RP+D+E  Q AG+ +  +NLR
Sbjct:  85  SSGFKRKPNPESMLYLREKYQISSGLV--IGDRPIDIEAGQAAGLDTHLFTSIVNLR 139

```

SEQ ID 8986 (GBS240) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 2; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 3; MW 51.5kDa).

30 GBS240-GST was purified as shown in Figure 225, lane 12.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2149

35 A DNA sequence (GBSx2265) was identified in *S.agalactiae* <SEQ ID 6641> which encodes the amino acid sequence <SEQ ID 6642>. Analysis of this protein sequence reveals the following:

```

   Possible site: 36
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.2620 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6643> which encodes the amino acid sequence <SEQ ID 6644>. Analysis of this protein sequence reveals the following:

```

   Possible site: 54
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.2967 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-2422-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 463/599 (77%), Positives = 541/599 (90%)

5 Query: 1 MSDNRSHIEEKYQWDLTVFATDELWETEVVELTQALDNAKGFSGHLLDSSQSLEITEV 60
M+DNRSH+EEKY WDL+T+T+T+ WE EV +L ++ +KGF+GHLLDSS +LL++T+
Sbjct: 1 MTDNRSHLEEKYTWDLSTIFATDKDWEAEVSDLA TEV EASKGFAGHLLDSSANLLKVKTKT 60

10 Query: 61 ELDLSRRLEKVVYVYASMKNDQD TT VAKYQE FQAKATALYAKFSETFSFYEP ELLQLSESD 120
L+L+RR+EKVVYVY MKNDQD TT VAKYQE+QAKA+ LYAKFSE FSFY+PE++ L + D
Sbjct: 61 YLELARRVEKVVYVYAHMKNDQD TT VAKYQEYQAKASGLYAKFSEVFSFYDPEVMMLHQED 120

15 Query: 121 YQSFLLLEMPDLQKYDHF FEKIFANKPHVLSQNEEELLAGASEIFGAAGETFEILDNADMV 180
YQ+FL E P+L+ Y+HFF+K+F + HVLSQ EEELLAGA EIF A ETF ILDNAD+V
Sbjct: 121 YQAFLTETPELKVYNHFFDKLFQAREHVLSQAE EELLAGAEIFNGABETFSILDNADIV 180

20 Query: 181 FPVVKNAKGEEVELTHGNFISLMESDR TVRKEAYQAMYSTYEQFOHTYAKTLQTNVKSQ 240
FPVVKN KGE+VELTHGNFISLMES DR+VR+ AY+AMYSTYEQFOHTYAKTLQTNV K Q
Sbjct: 181 FPVVKNDKGEDVELTHGNFISLMESKDRSVRQAAYEAMYSTYEQFOHTYAKTLQTNVKVQ 240

25 Query: 241 NFKARVHHYQSARQSALSANFIPEEVYETLIKT VNNHLLPLLHRYMKLRQKVLGLDDLKMY 300
N+KARVH Y SARQ+A++ANFIPE VY+TL++TVN HLP LLHRY+KLRQ+VLGLDDLKMY
Sbjct: 241 NYKARVHKYDSARQAAMAANFIPEAVYDTLLET VNNKHLPLLHRYLKL RQEV LGLDDLKMY 300

30 Query: 301 DVYTPLSQMDMSFTYDEAL KSEEVL AIFGEAYSERVHRAFTERWIDVHV NKGKRS GAYS 360
DVYTPLS+ D++ YDEAL+K+E+VLA+FG+ Y++RVHRAFTERWIDVHV NKGKRS GAYS
Sbjct: 301 DVYTPLSETDLAIGYDEALEKA EKVLAVFGKDYADRVHRAFTERWIDVHV NKGKRS GAYS 360

35 Query: 361 GGSYDTNAPMLLNWQDTLDNLYTLVHETGHS LHSFTFTRENQPYVYGDYSIFLAEIASTTN 420
GGSYDTNAP+LLNWQDTLDNLYTLVHETGHS LHSFTFTRE QPYVYGDYSIFLAEIASTTN
Sbjct: 361 GGSYDTNAPILLNWQDTLDNLYTLVHETGHS LHSFTFTRETQPYVYGDYSIFLAEIASTTN 420

40 Query: 421 ENILTETLLKEVKDDKNRFAILNHYLDGFGKGTIFRQTQFAEF EHAHIVADQEGQVLTSEY 480
ENI+TE LL EV+D+K RFAILNHYLDGFG+GT+FRQTQFAEF EHAH ADQ+G+VLTSEY
Sbjct: 421 ENIMTEALLNEVQDEKERFAILNHYLDGFRGT VFRQTQFAEF EHAHQADQKGEVLTSEY 480

45 Query: 481 LNNLYAELNEKYGLTKEDNHFIQYEWARIPHFYNNYVYQYATGF AAANYLAERIVNGN 540
LN LYA+LNEKYGL+K+DNHFIQYEWARIPHFYNNYV+QYATGF AAA+YLA++IV+G
Sbjct: 481 LNQLYADLNEKYGLSKDNHFIQYEWARIPHFYNNYVYQYATGF AAASYLADKIVHGT 540

Query: 541 PEDKEAYLNYLKAGNSDYPLNVIKAGVDMTSADYLDAAFRVFEERLVELENLVAKGVH 599
+D + YL YLK+GNSDYPL VIAKAGVDM DYL+AAF+VF+ERL ELE LV+KG+H
Sbjct: 541 QDDIDHYLAYLKSGNSDYPLNVIKAGVDM EKGDYLEAAFKVFDERLLEVLVSKGIH 599

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2150

A DNA sequence (GBSx2266) was identified in *S.agalactiae* <SEQ ID 6645> which encodes the amino acid sequence <SEQ ID 6646>. This protein is predicted to be competence protein. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2955(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the GENPEPT database.

-2423-

>GP:AAC23746 GB:AF052209 competence protein [Streptococcus pneumoniae]
Identities = 127/269 (47%), Positives = 176/269 (65%), Gaps = 8/269 (2%)

5 Query: 1 MLIADKQGNLINLLESHPGKGYFCPTCCSAVRLKAGRIMRRHFAHISLKNQCQFYHENE 60
M +A+D +G L+N+LE K Y CP C + L+ G +R HFAH SLK+C F+ ENE
Sbjct: 1 MFVARDARGELVNVLEDKLEKQAYTTPACGGQLHLRQGPSVRTHFAHKSLLKDCDFFENE 60

10 Query: 61 SNEHLQLKAKLYMSLSRENEMLEHHLPEINQIADLFVNETLALE----VQCSRLSEQRL 116
S EHL K LY L +E + LE+ L E+ QIAD+FVN LALE V C + + L
Sbjct: 61 SPEHLANKESLYHWLKKETKVQLEYPLSELKQIADVFVNGNLALESSVVVPCPK---KVL 117

15 Query: 117 RERTKAYLQADFQVRWLLGEKWLKHLRLTNLHKQFLQFSQSIGFHIWELDLRLEVLRLKY 176
+ER++ Y +QV WLLG+KLWLK RLT L FL FSQ++GF++WELD +VLRKY
Sbjct: 118 KERSEGYRSQGYQVLLGQKLWLKERLRLQAGFLYFSQNMGFYWELDKGKQVLRKY 177

Query: 177 LIYEDLRGHVYLSKTCPL-SGDVLAFLKWPYQSKNLNFKYKQDRNIRDYVRQQLRYGN 235
LIY+DLRG ++Y K G +L L+ PY+ + ++ + V +D++I Y+RQQL Y N
Sbjct: 178 LIYQDLRGKLVYQIKFESYQGSLLLEILRLPYKKQKISHFTVSEDKDICRYIRQQLYYQN 237

20 Query: 236 QFWLRKQEKAYLSGQNLLTQELMMFFPQI 264
FW+++Q +AY G+N+LT L ++PQI
Sbjct: 238 LFWMKQABAYQKGENILTYGLKEWYFQI 266

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6647> which encodes the amino acid
sequence <SEQ ID 6648>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1034(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 154/312 (49%), Positives = 204/312 (65%), Gaps = 1/312 (0%)

Query: 1 MLIADKQGNLINLL-ESHPGKGYFCPTCCSAVRLKAGRIMRRHFAHISLKNQCQFYHEN 59
+L A D + LI+L+ + K + CP C S VRL+ G I R HFAH+ L +CQF EN
40 Sbjct: 4 ILTALDDKNQLISLVTQPISTKPPRCPACKSPVRLRQGTIRPHFAHVQLAHCQFQAEN 63

Query: 60 ESNEHLQLKAKLYMSLSRENEMLEHHLPEINQIADLFVNETLALEVQCSRLSEQRLRER 119
ES EHL LKAKLY SL R +E +LPE+ QIADL+VN+ LALE+QCS L +RL++R
Sbjct: 64 ESEHLTLKAKLYTSLVRTEAVCIEKYLPELQIADLWVNDKLALAEIQCSPLPVERLKKR 123

45 Query: 120 TKAYLQADFQVRWLLGEKWLKHLRLTNLHKQFLQFSQSIGFHIWELDLRLEVLRLKYLIY 179
TKAY + + VRWLLG KLWL LT L KQFL FS S+GFH+WELD +LRLKYLI+
Sbjct: 124 TKAYQEKGYPVWRLLGRKLWLNTHLTALQKQFLYFSSSLGFHLWELDAAANLLRLKYLIH 183

50 Query: 180 EDLRGHVYLSKTCPLSGDVLAFLKWPYQSKNLNFKYKQDRNIRDYVRQQLRYGNQFWL 239
EDL G V YL+KT L +++ + PYQ + L Y+ K N+ +++ L + WL
Sbjct: 184 EDLFGKVSYLTKTISLDHNIMEMFRLPYQOEILYSYQKKMTVNLSKRIQRALLARHPKWL 243

55 Query: 240 RKQEKAYLSGQNLLTQELMMFFPQIQPPRVDTFQCITNSLTSFYQNFTNYYQKNKNND 299
R+QEKAYLSG NLL F+PQ +P + + FCQI +L +Y++F YY+K K+
Sbjct: 244 RRQEKAYLSGYNLLMTTDAFYQWRPVQSSSGFCQIKGNLRPYYESFKVYKKEKDKKV 303

Query: 300 QTLYPPVFYDKI 311
QTL+ P +Y K+
Sbjct: 304 QTLFSPKYVVKM 315

60

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2424-

Example 2151

A DNA sequence (GBSx2267) was identified in *S.galactiae* <SEQ ID 6649> which encodes the amino acid sequence <SEQ ID 6650>. This protein is predicted to be bicyclomycin resistance protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 25
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -8.33    Transmembrane    78 - 94 ( 75 - 96)
      INTEGRAL    Likelihood = -8.33    Transmembrane    269 - 285 ( 267 - 287)
      INTEGRAL    Likelihood = -7.38    Transmembrane    290 - 306 ( 287 - 314)
10   INTEGRAL    Likelihood = -7.06    Transmembrane    203 - 219 ( 199 - 225)
      INTEGRAL    Likelihood = -6.69    Transmembrane    157 - 173 ( 143 - 184)
      INTEGRAL    Likelihood = -6.42    Transmembrane    53 - 69 ( 44 - 73)
      INTEGRAL    Likelihood = -6.42    Transmembrane    362 - 378 ( 357 - 381)
      INTEGRAL    Likelihood = -3.72    Transmembrane    242 - 258 ( 240 - 261)
15   INTEGRAL    Likelihood = -3.24    Transmembrane    329 - 345 ( 328 - 346)
      INTEGRAL    Likelihood = -1.28    Transmembrane    107 - 123 ( 106 - 123)

    ----- Final Results -----
      bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
20   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

25   >GP:CAA15047 GB:AJ235272 BICYCLOMYCIN RESISTANCE PROTEIN (bcr1)
      [Rickettsia prowazekii]
      Identities = 86/336 (25%), Positives = 159/336 (46%), Gaps = 28/336 (8%)

Query: 73 GKKNTVLLGLCLILMSGFISFFTSNFSLAMASRLLLGIGIGLYNSLSISITDLYEADER 132
      G++ VLLGL + ++S IS F+ N + M +R + G+ + + + S+ D Y+ E
30   Sbjct: 70 GRRPIVLLGLFIYIVSSIISIFSFNIEMLMARFIQAFGVSVGSVIGQSMARDSYQGAEL 129

Query: 133 ASMIGLRTASLNIGKALTTFIVGLVLA-IGVNYIYLVYLLVIPVF-FFWKNVPEVENQT 190
      + + + + L AL ++I G ++ + +Y+++ + L + +++ +PE
35   Sbjct: 130 SYVYAILSPWLLFIPALGSGYIGGYIIEYLSWHYVFIFFSLAGTILLALYYQILPETNYII 189

Query: 191 HTLKASTTFDT-----KAALLMLITFLVGI---AYIGATVKIPTLLVTKYHYATSFSSNM 242
      ++S F+ K +L L F++G Y G ++ P +L+ + SF +
40   Sbjct: 190 AFSQSSKYFEVFNIIKDKMLWLYAFIIGAFNGIYYGFIEAPFILIDQMRVLPSPFYKGL 249

Query: 243 LTLAFSGILVGSVFGKLVK---VFQEKTLIMILAMGIGNVLFALANNQIIFIVAS--I 297
      LL+F+ I G + G L+K V+ +K + I + G +LFA+ + + FI+ S
45   Sbjct: 250 AFLLSFASIFGGFLGGYLIKKRQVYDKKMSIGFIFSLCGCILFAVDSFILEFILVSNVF 309

Query: 298 LIGASFVGTM-----SSVFFYISKNYAKEHNNFITSLLTAGNI-GVILTPLI--LTKLP 349
      I F+ M S+ I+ YA E +T TAG+I G I +I +T
50   Sbjct: 310 AIAMIFMPMMIHMIGHSLIIAITLRYALEDYATVTG---TAGSIFGAIYYVVIASVTYCV 366

Query: 350 SOLHLEPFMTPLITSGLMVINV--FVYLVLMSSNK 383
      S++H E L+ L + +V F Y+ L+ K K
      Sbjct: 367 SKIHGETISNFSLLCLVLSISSVISFYIICLLYKKK 402

```

A related GBS gene <SEQ ID 8987> and protein <SEQ ID 8988> were also identified. Analysis of this protein sequence reveals the following:

```

55   Lipop: Possible site: -1    Crend: 7
      McG: Discrim Score: 6.28
      GvH: Signal Score (-7.5): -2.45
      Possible site: 25
      >>> Seems to have a cleavable N-term signal seq.
      ALOM program count: 10 value: -8.33 threshold: 0.0
60   INTEGRAL    Likelihood = -8.33    Transmembrane    78 - 94 ( 75 - 96)
      INTEGRAL    Likelihood = -8.33    Transmembrane    269 - 285 ( 267 - 287)
      INTEGRAL    Likelihood = -7.38    Transmembrane    290 - 306 ( 287 - 314)

```


-2426-

Example 2152

A DNA sequence (GBSx2268) was identified in *S.agalactiae* <SEQ ID 6651> which encodes the amino acid sequence <SEQ ID 6652>. This protein is predicted to be 16S pseudouridylate synthase (rsuA). Analysis of this protein sequence reveals the following:

5 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2645(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:BAB06992 GB:AP001518 16S pseudouridylate synthase [Bacillus halodurans]
 Identities = 106/234 (45%), Positives = 141/234 (59%), Gaps = 1/234 (0%)

 Query: 1 MRLDKLLGQAGFGSRNQVKKLCISRQVSVDDGQIVTKDNVIVDSGLQSFVVGKERVCLKES 60
 MR+DK L GFGSR VKKL+ + V V GQ + + V+ +SI V E V K
 Sbjct: 1 MRIDKFLANMGFGSRKDVKKLLKTGAVRVQGQPIKDPSTHVPESESITVYGEEVEYKPY 60

20 Query: 61 SYLLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPIGRLDRDTEGLLIVTNNGLPGY 120
 Y ++ KP GV+ A D EH+TVIDL+ E+++ P+GRLD+DT GLL++TN+G +
 Sbjct: 61 VYLMMNKPKGVICATEDLEHETVIDLLGEEERHYEPSPVGRLDKDTVGLLLITNDGKFNH 120

25 Query: 121 RMLHPKHHVAKTYYYVEVNGFLERDAITFFEEGVVFDGDKCKPAELTIDTANNDKSTARI 180
 ++ PKHHV KTY V G + + + F GVV DDG KPA L I A +S +
 Sbjct: 121 WLMSPKHHVPKTYRALVEGHVTEEDVGAFSHGVVLDGQVYTKPATLHILEA-GARSHIEL 179

30 Query: 181 TITEGKFHQVKMFAYGVKVIYLRISFGDLRLDMNLKPGQYRRLDSEAAIL 234
 +TEGKFHQVK+MF A G +V+ L RI G+L LD L G+YR L E A+L
 Sbjct: 180 ILTEGKFHQVKRMFQAVGKRVLELERIKIGNLLLDPELARGEYRELTKETIAL 233

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6653> which encodes the amino acid sequence <SEQ ID 6654>. Analysis of this protein sequence reveals the following:

35 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3310(Affirmative) < succ>
 40 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 111/194 (57%), Positives = 138/194 (70%)

 Query: 1 MRLDKLLGQAGFGSRNQVKKLCISRQVSVDDGQIVTKDNVIVDSGLQSFVVGKERVCLKES 60
 MRLDKLL GSR+QVKKLI ++ V VD VD GLQ I V +RV +
 Sbjct: 1 MRLDKLLEGTQVGSRSQVKKLIKAQGVVVDHMPARNGRQNVDPGLQLIEVTGQRVTHPKH 60

50 Query: 61 SYLLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPIGRLDRDTEGLLIVTNNGLPGY 120
 SY +L KPSGVVSA +D+ + TVID ++E+DK LYP+GRLD+DTGL+++T+NGPLG+
 Sbjct: 61 SYIILNKPSGVVSAKRDNTYLTVIDQLAEEDKSPDLYPVGRLD+DTGLVLLTDNGPLGF 120

55 Query: 121 RMLHPKHHVAKTYYYVEVNGFLERDAITFFEEGVVFDGDKCKPAELTIDTANNDKSTARI 180
 RMLHP HHV+KTY V VNG L DA FF G+ F G +C+PA+LTI A+ D+S A +
 Sbjct: 121 RMLHPSHHVSKTYLVTVNGLLAEDASDFFAAGICFPTGEQCQPAQLTILKADTDQSQASL 180

60 Query: 181 TITEGKFHQVKMF 194
 TI+EGKFHQVKK F
 Sbjct: 181 TISEGKFHQVKKCF 194

-2427-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2153

A DNA sequence (GBSx2269) was identified in *S.agalactiae* <SEQ ID 6655> which encodes the amino acid sequence <SEQ ID 6656>. Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9745> which encodes amino acid sequence <SEQ ID 9746> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:BAA18872 GB:D90917 hypothetical protein [Synechocystis sp.]
Identities = 197/318 (61%), Positives = 243/318 (75%)

20 Query: 22 MGLLVDPKQWVDQWYDTASTGGKFVRTVTQFRHWVTKDGSAGPSGDAGFKAESGRYHLYVS 81
 MGLLV+G W DQWYDT STGG+FVR +QFRHW+T DGS GP+G GFKA+GRYHLYVS
Sbjct: 1 MGLLVNGIQDQWYDTESTGGRFVRHDSQFRHWITPDGSPGPTGHGGFKAAGRYHLYVS 60

25 Query: 82 LACPWASRVLMRKLNLESHISISIVNPLMLENGWTFQFYKGVIPDMINQSQYLYQIYQ 141
 LACPWA R LI RKLK LE I +S+V+ LM ENGWTF GV+PD + ++YLYQIY
Sbjct: 61 LACPWAHRTLIFRKLKGLEGMIDVSVVHWMRENGWTFAPGPGVMPDPLFNAEYLYQIYT 120

30 Query: 142 ASQSDYTGRVTVPVLWDKKFHTIVNNESEIMRMLNTAFNHITGNTDDYYPDSLQGOIDE 201
 + + Y+GRVTVP+LWDK+ TIVNNESEI+R+ N+AF+ + + DYYP +L+ QID
Sbjct: 121 RADAQYSGRVTVPILWDKQKQTIIVNNESEIIRIFNSAFDGLGAKSGDYYPKALRTQIDA 180

35 Query: 202 MNNFIYPKINNGVYKAGFATSQNVYQKEVETLFTALDQLEKHLSDNHLYVGEQFTEADIR 261
 +N+ IY INNGVYK GFAT+Q Y++ + LF +LD LE L + YL G++ TEAD R
Sbjct: 181 LNDRIYHTINNGVYKCGFATTQTAYEEAIAPLFESLDWLEGILQGHQYLTGDEITEADWR 240

40 Query: 262 LFTTLVRFDTVYVGHEFKCNLKHLDYPHLWHYTKRIYNLPGIAETVNFHDIKKHYGSHK 321
 LFTTL+RFD VY GHEFKNL+ + DYP+LW Y + +Y+ PGIAETVNF HIK HYY SH
Sbjct: 241 LFTTLIRFDVVYVGHEFKCNLRRIQDYPNLWRYLRDLYHQPGIAETVNFQHIKHYYESHL 300

40 Query: 322 TINPTGIIPAGPNLDWTI 339
 INPTGI+P GP LD ++
Sbjct: 301 NINPTGIVPMGPALDLSL 318

No corresponding DNA sequence was identified in *S.pyogenes*.

45 SEQ ID 6656 (GBS655) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 2-4; MW 27kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2154

50 A DNA sequence (GBSx2270) was identified in *S.agalactiae* <SEQ ID 6657> which encodes the amino acid sequence <SEQ ID 6658>. Analysis of this protein sequence reveals the following:

Possible site: 43

-2428-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1116(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase
 [Bacillus subtilis]
 Identities = 112/243 (46%), Positives = 163/243 (66%), Gaps = 10/243 (4%)

15 Query: 1 MRVITVKNIDIEGGKIAFTLLLEKMKAGAQT-LGLATGSSPITFYEEIVKS----NLDFSN 55
 M+++ ++ E K++ +++E+++A LGLATGS+P+ Y++++ +DFS
 Sbjct: 1 MKILIAEHYEELCKLSAAIIKEBQIQAKDAVLGLATGSTPVGLYKQLISDYQAGEIDFSK 60

20 Query: 56 MVSINLDEYVGLIASNDQSYSYFMHKLFDAPKPFKENNL--PNGLAKDLKEEIKRYDAVI 113
 + + NLDEY G++ S+ QSY++FMH+HLF + +++ P G L+ K Y+ +I
 Sbjct: 61 VTTFNLDEYAGLSPSHPQSYNHFMEHLFQHINMQPDHIHIPQGDNPQLEAACKVYEDLI 120

25 Query: 114 N-ANPIDFQILGIGRNGHIGFNEPGTPFDITTHVVDLAPSTIEANSRFFNSIDD-VPKQA 171
 A ID QILGIG NGHIGFNEPG+ F+ T VV L+ STI+AN+RFF VP+ A
 Sbjct: 121 RQAGGIDVQILGIGANGHIGFNEPGSDFEDRTRVVKLSESTIQANARFFGGDPVLVPRLA 180

30 Query: 172 LSMGIGSIMK-SKTIVLVAYGIEKAEAIASMIKGPITEDMPASILQKHDDVVIIVDEAAA 230
 +SMGI +IM+ SK IVL+A G EKA+AI M +GP+T D+PASILQKH+ V +I D AA
 Sbjct: 181 ISMGIKTIMEFSKHIVLLASGEEKADAIQKMAEGPVTDDVPASILQKHNVTVIADYKAA 240

35 Query: 231 SKL 233
 KL
 Sbjct: 241 QKL 243

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6659> which encodes the amino acid sequence <SEQ ID 6660>. Analysis of this protein sequence reveals the following:

35 Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 174 - 190 (174 - 190)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase
 [Bacillus subtilis]
 Identities = 120/244 (49%), Positives = 162/244 (66%), Gaps = 12/244 (4%)

50 Query: 1 MKIIRVQDQIEGGKIAFTLLKDSL-AGKAKTLGLATGSSPISFYQEMVKS----PLDFSD 55
 MKI+ + E K++ ++K+ + AK LGLATGS+P+ Y++++ +DFS
 Sbjct: 1 MKILIAEHYEELCKLSAAIIKEBQIQAKDAVLGLATGSTPVGLYKQLISDYQAGEIDFSK 60

55 Query: 56 LTSINLDEYVGLSVESDQSYDYFMRQNLF---NAKPFKKNYLPNGLATDVEAEAKRYNQI 112
 +T+ NLDEY GLS QSY++FM ++LF N +P ++P G +EA K Y +
 Sbjct: 61 VTTFNLDEYAGLSPSHPQSYNHFMEHLFQHINMQP-DHIHIPQGDNPQLEAACKVYEDL 119

60 Query: 113 IAEHP-IDFQVLGIGRNGHIGFNEPGTSFEEETHVVDLQESTIEANSRFFTSIED-VPKQ 170
 I + ID Q+LGIG NGHIGFNEPG+ FE+ T VV L ESTI+AN+RFF VP+
 Sbjct: 120 IROAGGIDVQILGIGANGHIGFNEPGSDFEDRTRVVKLSESTIQANARFFGGDPVLVPRIL 179

 Query: 171 AISMGIASIMK-SEMIIVLAFQEKADAIKGMVFGPITEHLPASILQKHDHVIVIVDEAA 229
 AISMGI +IM+ S+ IVLLA G+EKADAI+ M GP+T +PASILQKH+HV VI D A
 Sbjct: 180 AISMGIKTIMEFSKHIVLLASGEEKADAIQKMAEGPVTDDVPASILQKHNVTVIADYKA 239

-2429-

Query: 230 ASQL 233

A +L

Sbjct: 240 AQKL 243

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 163/233 (69%), Positives = 201/233 (85%)

Query: 1 MRVITVKNIDIEGGKIAFTLLLEEKMKAGAQTGLLATGSSPITFYEEIVKSNLDFSNMVSIN 60
 M++I V++ IEGGKIAFTLL++ + GA+TLGLATGSSPI+FY+E+VKS LDFS++ SIN
 10 Sbjct: 1 MKIIRVQDQIEGGKIAFTLLKDSLAKGAKTLGLATGSSPISFYQEMVKSPLDLFSDLTSIN 60

Query: 61 LDEYVGIAASNDQSYSYFMHKHLFDAKPFKENNLPNGLAKDLKEEIKRYDAVINANPIDF 120
 LDEYVG++ +DQSY YFM ++LF+AKPFK+N LPNGLA D++ E KRY+ +I +PIDF
 15 Sbjct: 61 LDEYVGLSVESDQSYDYFMRQNLFNAPFKKNYLPNGLATDVEAEAKRYNQIIAEHPIDF 120

Query: 121 QILGIGRNGHIGFNEPGTFFDITTHVVDLAPSTIEANSRFFNSIDDPKQALSMGIGSIM 180
 Q+LGIGRNGHIGFNEPGT F+ THVVDL STIEANSRFF SI+DVPKQA+SMGI SIM
 Sbjct: 121 QVLGIGRNGHIGFNEPGTSFEEETHVVDLQESTIEANSRFFTSIEDVPKQAISMGIASIM 180

Query: 181 KSKTIVLVAYGIEKAEAIASMIKGPITEDMPASILQKHDDVVIIVDEAAASKL 233
 KS+ IVL+A+G EKA+AI M+ GPITE +PASILQKHD V++IVDEAAAS+L
 20 Sbjct: 181 KSEMIVLLAFGQEKADAIKGMVFGPITEHLPASILQKHDHVIVIVDEAAASQL 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 2155

A DNA sequence (GBSx2271) was identified in *S.agalactiae* <SEQ ID 6661> which encodes the amino acid sequence <SEQ ID 6662>. Analysis of this protein sequence reveals the following:

Possible site: 61
 30 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.12	Transmembrane	169 - 185 (161 - 194)
INTEGRAL	Likelihood = -6.37	Transmembrane	151 - 167 (145 - 168)
INTEGRAL	Likelihood = -5.15	Transmembrane	42 - 58 (41 - 62)
INTEGRAL	Likelihood = -1.59	Transmembrane	207 - 223 (207 - 224)
INTEGRAL	Likelihood = -1.12	Transmembrane	24 - 40 (23 - 40)

35 ----- Final Results -----

bacterial membrane	---	Certainty=0.4248 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF13747 GB:AF117351 unknown [Zymomonas mobilis]
 Identities = 88/216 (40%), Positives = 123/216 (56%)
 45

Query: 9 QQLNILRAGVLGANDGIISVAGVIGVASATHNLWIIFLSAASAILAGAFSMAGGEYVSV 68
 +Q+ LRA VLGANDGI+S + ++IGVASA + I L+ S ++AGA SMA GEYVSV
 Sbjct: 17 RQMCWLRASVLGANDGILSTSSLMIGVASAHGSSGNILLAGMSGIAGALSMAAGEYVSV 76

50

Query: 69 STQKQTEQAAVAREEKLLENPELAKKSLVDIYLAKGESHEHAQWLVDKAFSKNAIEHLV 128
 S+Q D EQA VARE L+ NP K L +IY+ +G E A + ++ + NA+E +
 Sbjct: 77 SSQHDMEQADVAREHAELKANPHEKHELAETIYVERGLDRELALQVAEQLMANALEAHL 136

55

Query: 129 EEKYGIEFGEYTSPWHAATSSFIAFAIGSIFPTITILLPFSVRIVGTVIIIVISLLSTG 188
 ++ G+ P AA++S I+F+ G+I P +T L P + + +I I+ L G
 Sbjct: 137 RDELGLTDSLIRPVQAALASAISSFSGGAIVPFLTALFSPPEIINITISLISILCLAVLG 196

60

Query: 189 YVSAKLGQAPTVPAMRRNMIGCLTMLATYVIGQLF 224
 V A LG A A R G L M+ T IG F
 Sbjct: 197 MVGAHLGGANVPKAALRVTFGALAMIGTAAIGSFF 232

-2430-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2156

- 5 A DNA sequence (GBSx2272) was identified in *Sagalactiae* <SEQ ID 6663> which encodes the amino acid sequence <SEQ ID 6664>. This protein is predicted to be S-adenosylmethionine tRNA ribosyltransferase (queA). Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

10

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3438(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14732 GB:Z99118 S-adenosylmethionine tRNA ribosyltransferase
[Bacillus subtilis]

Identities = 228/341 (66%), Positives = 279/341 (80%)

20

Query: 1 MNTNDFDFYLPEELIAQTPLEKRDASKLLVIDHKNTMTDSHFDHILDELKPGDALVMNN 60

M + FDF LPE LIAQ PLE+RDAS+L+V+D +TDS F HI+ GD LV+NN

Sbjct: 1 MKVDLDFDFELPERLIAQVPLEQRDASRLMVLDKHTGELTDSSEFKHIIISFNEGDCLVLNN 60

25

Query: 61 TRVLPARLYGEKQDTHGHVELLLKNTTEGDQWEVLAKPAKRLRVGTVKVSFGDGRLIATVT 120

TRVLPARL+G K+DT VELLLLK GD+WE LAKPAKR++ GT V+FGDGRL A T

Sbjct: 61 TRVLPARLFGTKEDTGAKVELLLKQETGDKWETLAKPAKRVKGTIVTFGDGRLKAICT 120

30

Query: 121 KELEHGGRIVEFSYDGFLEVLSELGEMPLPPYIHEKLEDRDRYQIVYAKENGSAAPTA 180

+ELEHGGR +EF YDGIF EVLESLGEMPLPPYI E+L+D++RYQTVY+KE GSAAAPTA

Sbjct: 121 EELEHGGRKMEFYDGFYFVLESLGEMPLPPYIKEQLDDKERYQTVYSKEIGSAAPTA 180

35

Query: 181 GLHFTKELLEKIETKGVKLVLTLHVGLGTFRPVSVNLDHEMHSEFYQLSKEAADTLN 240

GLHFT+E+L+++ KGV++ ++TLHVGLGTFRPVS D ++EH MH+EFYQ+S+E A LN

Sbjct: 181 GLHFTTEILQQLKDKGVQIEFITLHVGLGTFRPVSADEVEEHNMAEFYQMSEETAALN 240

40

Query: 241 AVKESGGRIVAVGTTSTRTLETIGSKFNGELKADSGWTNIFIKPGYQFKVVDASTNFHL 300

V+E+CGRI++VGTTSTRTLETI + +G+ KA SGWT+IFI PGY+FK +D TNFHL

Sbjct: 241 KVRENGGRIISVGTTSTRTLETIAGEHDGQFKASSGWTISIFIYPGYBFKAIDGMITNFHL 300

Query: 301 PKSTLVMLVSAFAGRDFVLEAYNHAVEERYRFFSFGDAMFV 341

PKS+L+MLVSA AGR+ +L AYNHAVEE YRFFSFGDAM +

Sbjct: 301 PKSSLIMLVSAAGRENILRAYNHAVEEYRFFSFGDAMLI 341

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6665> which encodes the amino acid sequence <SEQ ID 6666>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3864(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 55 An alignment of the GAS and GBS proteins is shown below.

Identities = 297/341 (87%), Positives = 322/341 (94%)

-2431-

Query: 1 MNTNDFDFYLPPELIAQTPLEKRDASKLLVIDHKNKMTDSHFDHILDELKPGDALVMNN 60
 MNTN+FD F LPEELIAQTPLEKRD+SKLL+IDH+ KTM DSHFDHI+D+L PGDALVMNN
 Sbjct: 1 MNTNDFDFELPEELIAQTPLEKRDSSKLLIIDHRQKTMVDSHFDHIIDQLNPGDALVMNN 60

5 Query: 61 TRVLPARLYGEKQDTHGHVELLLLKNTGEGDQWEVLAKPAKRLRVGTVKVSFGDGRLIATVT 120
 TRVLPARLYGEK DTHGHVELLLLKNT+GDQWEVLAKPAKRL+VG++V+FGDGR L AT+
 Sbjct: 61 TRVLPARLYGEKPDTHGHVELLLLKNTQGDQWEVLAKPAKRLKVG SQVNFGDGR LKATII 120

10 Query: 121 KELEHGGRIVEFSYDGIFLEVLES LGEMPLPPYIHEKLED RDRYQT VYAKENGSAAPTA 180
 ELEHGGRIVEFSYDGIFLEVLES LGEMPLPPYIHEKLED +RYQT VYAKENGSAAPTA
 Sbjct: 121 DELEHGGRIVEFSYDGIFLEVLES LGEMPLPPYIHEKLEDAERYQT VYAKENGSAAPTA 180

15 Query: 181 GLHFTKELLEKIE TKGVKL VYLT LHVGLGTFRPVSV DNLDEHMHSEFY QLSKEADTIN 240
 GLHFT +LL+KIE KGV LVYLT LHVGLGTFRPVSV DNLDEH+MHSEFY LS+EAA TL
 Sbjct: 181 GLHFTD LLLKIEAKGVHL VYLT LHVGLGTFRPVSV DNLDEHDMHSEFYSLSEEAQT LR 240

20 Query: 241 AVKESGGRIVAVGTT SIRTLETIGSKFNGELKADSGWTNIFIKPGYQFKVVD AFS TNFHL 300
 VK++GGR+VAVGTT SIRTLETIG K F G+++ADSGWTNIFIKPGYQFKVVD AFS TNFHL
 Sbjct: 241 DVKQAGGRVAVGTT SIRTLETIGGKFQGD IQADSGWTNIFIKPGYQFKVVD AFS TNFHL 300

Query: 301 PKSTLVMLVSAFAGRDFVLEAYNHAVEERYRFFSFGDAMFV 341
 PKSTLVMLVSAFAGRDFVLEAY HAV+E+YRFFSFGDAMFV
 Sbjct: 301 PKSTLVMLVSAFAGRDFVLEAYRHA VDEKYRFFSFGDAMFV 341

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2157

A DNA sequence (GBSx2273) was identified in *S.galactiae* <SEQ ID 6667> which encodes the amino acid sequence <SEQ ID 6668>. Analysis of this protein sequence reveals the following:

30 Possible site: 36
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -14.22 Transmembrane 14 - 30 (6 - 34)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6669> which encodes the amino acid sequence <SEQ ID 6670>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2655(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 126/195 (64%), Positives = 155/195 (78%), Gaps = 1/195 (0%)

Query: 160 MEERFDITETDY EYIGEHN NYVAAFSGAMSIDDMQKYS LVYSENTPAYALAERIGGMDSA 219
 M ERFDITETDY EY EH+ YVA F+GAMSI DMQ+YSLVYSENTPAYALAER+GGM+ A
 55 Sbjct: 1 MTERFDITETDY EYDQEH HAYVAQFNGAMSIPDMQ EYSLVYSENTPAYALAERLGGMNKA 60

Query: 220 YSKFGRYQSGKGD IKNIQKNGNKVTTDY YIQVLDYLWKHKYDSLITYLEEAFPTDYR 279
 Y F RYG+ G I I +NGNK+TT YY+QVLDYLW+H+ KY ++ Y+ E+FP YY+
 Sbjct: 61 YQLFDRYGVSGAITTIDRNGNKITAYYLQVLDYLWQH QDKYKDILYYIGESFPDLYYK 120

-2432-

Query: 280 ALIPSDVVVAQKPGYVREALNVGAIVKEEVPIVAIYTAGLGGSTQEDSEINGVGLYQLE 339
 +P V V QKPGYVREALNVGAIV EE PY++A+Y++GLGG+TQ E+NG+G QL
 Sbjct: 121 TYLP-HVKVYQKPGYVREALNVGAIVCEESEPYLIALYSSGLGGATQASEEVNGLGYQLV 179

Query: 340 QLCFVINQWHRVNMN 354
 QL +VIN+W+R N+N
 Sbjct: 180 QLPYVINEWYRGNLN 194

SEQ ID 6668 (GBS680) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 10-12; MW 64kDa) and in Figure 239 (lane 9; MW 64 kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 15; MW 40kDa) and in Figure 188 (lane 9; MW 40kDa). Purified GBS680-His is shown in Figure 242, lane 8. Purified GBS680-GST is shown in Figure 246, lanes 6 & 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2158

A DNA sequence (GBSx2274) was identified in *S.galactiae* <SEQ ID 6671> which encodes the amino acid sequence <SEQ ID 6672>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -4.57	Transmembrane	8 - 24 (4 - 25)
INTEGRAL	Likelihood = -2.13	Transmembrane	66 - 82 (65 - 84)
INTEGRAL	Likelihood = -1.65	Transmembrane	107 - 123 (107 - 125)
INTEGRAL	Likelihood = -0.69	Transmembrane	36 - 52 (36 - 52)
INTEGRAL	Likelihood = -0.48	Transmembrane	89 - 105 (89 - 105)

----- Final Results -----

bacterial membrane	---	Certainty=0.2826(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2159

A DNA sequence (GBSx2275) was identified in *S.galactiae* <SEQ ID 6673> which encodes the amino acid sequence <SEQ ID 6674>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.87	Transmembrane	108 - 124 (97 - 133)
INTEGRAL	Likelihood = -9.08	Transmembrane	181 - 197 (173 - 201)
INTEGRAL	Likelihood = -7.43	Transmembrane	220 - 236 (216 - 248)
INTEGRAL	Likelihood = -6.69	Transmembrane	6 - 22 (3 - 28)
INTEGRAL	Likelihood = -3.72	Transmembrane	401 - 417 (400 - 417)
INTEGRAL	Likelihood = -3.35	Transmembrane	279 - 295 (278 - 295)
INTEGRAL	Likelihood = -2.87	Transmembrane	31 - 47 (30 - 50)
INTEGRAL	Likelihood = -2.87	Transmembrane	244 - 260 (242 - 264)
INTEGRAL	Likelihood = -0.80	Transmembrane	62 - 78 (62 - 78)

----- Final Results -----

-2433-

bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21770 GB:U32694 H. influenzae predicted coding region HI0092
 [Haemophilus influenzae Rd]
 Identities = 232/416 (55%), Positives = 314/416 (74%), Gaps = 3/416 (0%)

10 Query: 4 TFTTTGALIGLALAILLIKKVHPAYSILILGALVGGGLIGGDLVTIVNTMVLGAQGMSS 63
 T + GAL+ L +AI LI+KKV PAY +++GALVGGGLIGG DL V+ M+ GAQG+ ++
 Sbjct: 3 TVSAIGALVALIVAIIFLILKKVSPAYGMLVGVGGGLIGGADLSQTVSLMIGGAQGITTA 62

15 Query: 64 ILRILTSGLILAGALIKTGSAEKIAESIIKKLQQRITALAIATMIICAVGVFIDIAVIT 123
 ++RIL +G+LAG LI++G+A I E+I KLG+ RA+ ALA+ATMI+ AVGVF+D+AVIT
 Sbjct: 63 VMRILAAGVLAGVLIESGAANSITETITTNKLGETRALLALALATMILTAVGVFVDVAVIT 122

20 Query: 124 VAPIALAIGKKANLSKSSILLAMIGGGKAGNIISPNTIAASEAFKVDLTSLMVQNIIP 183
 V+PIALA+ ++++LSK++ILLAMIGGGKAGNI+SPNPN IAA++ F + LTS+M+ IIP
 Sbjct: 123 VSPIALALSRRSDLSKAAILLAMIGGGKAGNIMSPNPNIAAADTFHLEPLTSVMAGIIP 182

25 Query: 184 AIAALVVTIILAKIVSKKNNDISYDSEEQV--GSDLEAFLEPAISGPLVVICLLALRPLFG 241
 A+ L++T LAK + K + ++ D E V +LP+FL A+ PLV I LLALRPLF
 Sbjct: 183 ALFGLILTYFLAKRLINKGSKVT-DKEVIVLETQNLPSFLTALVAPLVAILLLALRPLFD 241

30 Query: 242 ITIDPLIALPLGLISILATGYLKETVPFVEYGLSKVVGVSILLIGTGTLSGIKASNLQ 301
 I +DPLIALPLGLI G L+ + GLSK+ V+I+L+GTG L+GII S L+
 Sbjct: 242 IKVDPLIALPLGLIGAFCMGKLNRNINSYAINGLSKMTFVAIMLLGTGALAGIIANSGLK 301

35 Query: 302 FDMIHLLLEFLNMFTEFILAPLSGIFMGAATASTTSGTTIASQTFAETLIKSGVPAVSGAAM 361
 +I LE +P++ILAP+SG+ M ATASTT+GT +AS F+ TL++ GV +++GAAM
 Sbjct: 302 EVLIQGLEHSGLEPSYILAPISGVLMSLATASTTAGTAVASNVFSTLELGVSSLAGAAM 361

Query: 362 IHAGATVLDLPHGSFFHATGGAVNMAIKDRMKLISYEALIGLTSTIVAVVYCYFF 417
 IHAGATV D +PHGSFFHATGG+VNM IK+R+KLI YE+ +GL TIV+ + + F
 Sbjct: 362 IHAGATVFDHMPHGSFFHATGGSVNMDIKERLKLIPYSAVGLMMTIVSTLIFGVF 417

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6675> which encodes the amino acid sequence <SEQ ID 6676>. Analysis of this protein sequence reveals the following:

40 Possible site: 51
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.15	Transmembrane	240 - 256 (236 - 265)
INTEGRAL	Likelihood = -10.88	Transmembrane	3 - 19 (1 - 32)
INTEGRAL	Likelihood = -10.14	Transmembrane	269 - 285 (263 - 289)
45 INTEGRAL	Likelihood = -7.27	Transmembrane	107 - 123 (102 - 141)
INTEGRAL	Likelihood = -7.17	Transmembrane	307 - 323 (303 - 330)
INTEGRAL	Likelihood = -6.64	Transmembrane	24 - 40 (23 - 43)
INTEGRAL	Likelihood = -5.63	Transmembrane	422 - 438 (420 - 442)
INTEGRAL	Likelihood = -3.77	Transmembrane	124 - 140 (124 - 141)
50 INTEGRAL	Likelihood = -3.24	Transmembrane	189 - 205 (184 - 207)
INTEGRAL	Likelihood = -2.60	Transmembrane	65 - 81 (65 - 82)
INTEGRAL	Likelihood = -2.34	Transmembrane	393 - 409 (393 - 409)
INTEGRAL	Likelihood = -0.11	Transmembrane	149 - 165 (149 - 165)

55 ----- Final Results -----

bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

>GP:BAB07616 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 155/435 (35%), Positives = 248/435 (56%), Gaps = 21/435 (4%)

Query: 7 LGVLGVIVIIYLYVKEVNIIIAAPLATSLVILFNQMDPTTLLGKEPNQFMGALSTYIL 66

-2434-

LG+++G++++ L + +II AP+A +V LF +D LL + +M +
 Sbjct: 2 LGIVLGLVILMVLAIRGWSIIWVAPIAAGVVALFGGLD---LLPAYTDTYMEGFVNFAK 57
 Query: 67 NYFAIFLLGSILAKLMETSGATTSIADYILKKVGHDSPLYKVLVAIFLISAILTYGGISLF 126
 +F +F+LG+I KLME +GA S+A I K +G + ++ + L A+LTYGGISLF
 Sbjct: 58 QWFPVFMGLAIFGKLMEDTGAARSVASAITKLIGTK--RAILGVMLGCAVLTYGGISLF 114
 Query: 127 VVMFAVLPLARSLFKKMDLAWNLIQVPLWLGIATFTMTILPGTPAIQNVIPQYLDTSLT 186
 VV+FA+ PLA +LF++ +++ LI + LG TFTMT +PGTP IQN+IP Y T+
 Sbjct: 115 VVVFAMYPLALALFREANISRRLIPGTIALGAFTFTMTAVPGTPQIQNLIPTSYYGNTAM 174
 Query: 187 AAAIPSIIVGSIGCVAFGLFYMKYCLAKSMARGETYATYAFDNEIQVKTKNLPHFLASILP 246
 AA + ++ ++ G Y+ + K GE + T + E + + + +P+ S LP
 Sbjct: 175 AAPMMGVIAALIMGIGGYTTLVWREKKLKEAGE-FFTEPKNGEKEEKEKVPNPWLSFLP 233
 Query: 247 LLLLLIIIALTGSLFGNDFFKKNIIFIALLAVILTASWLFQFIPNKIAVFNLGASSSIAP 306
 L+ +I+ T +L D I +AL++ I+ L + I N GA S+
 Sbjct: 234 LVSVIV---TLNLLQWD-----IVLALISGIVLIMLLNVGKVKGFQIQSMNQAGGGSVLA 284
 Query: 307 IFATASAVAFGAVVMIVPGFTFFSDLIILNIPGNPLISLAVLTSSMSAITGSSSGALGIVM 366
 I T++AV FG+VV VPGF ++L+L I G+PLIS AV + ++ TGS+SG +GI +
 Sbjct: 285 IINTSAAVGFSGSVVRVPGFERLTELLELGIQGSPLISQAVAINVLGATGSASGGMGIAL 344
 Query: 367 ----PNFAQYYLDQGLNPMEIHRVATIASNIFTIVQSGVFLTFLALTGLNHKNFAKETF 422
 + Q ++ G++PE HRVA+IAS +P +G LT LA+TGL+HK ++K+ F
 Sbjct: 345 EALGDRYMLAMETGMSPEAFHRVASIASGGLDTLPHNGAVLTLAITGLSHKESYKDIF 404
 Query: 423 ITVSVSTFIAQVIVI 437
 + V ++ I
 Sbjct: 405 VVGCVIPIVSVAFAI 419
 An alignment of the GAS and GBS proteins is shown below.
 Identities = 88/395 (22%), Positives = 167/395 (42%), Gaps = 40/395 (10%)
 Query: 9 GALIGLALAILLIKKVHPAYSILGALVGGLIGGGDLVTIV---NTMVLGAQG--MMS 62
 G L+G+ + I L +K+V+ + L + L D T + +GA +++
 Sbjct: 8 GVLGVIVIIYLVVKEVNIIIAAPLATSILVILFNQMDPTTTLGKEPNQFMGALSTYILN 67
 Query: 63 SILRILTSGLAGALIKTGSAEKIAESIKKLGQQ---RAITALAIATMIICAVGVFIDI 119
 L ILA + +G+ IA+ I+KK+G + + A+ + + I+ G+ + +
 Sbjct: 68 YFAIFLLGSILAKLMETSGATTSIADYILKKVGHDSPLYKVLVAIFLISAILTYGGISLFV 127
 Query: 120 AVITVAPIALAIGKKANLSKSSILLAMIGGGKAGNII---SPNPNTIAASEAFKVDLTS 175
 + V P+A ++ KK +L+ + I + + G + +P + + LT+
 Sbjct: 128 VMFAVLPLARSLFKKMDLAWNLIQVPLWLGIATFTMTILPGTPAIQNVIPQYLDTSITA 187
 Query: 176 LMVQNIIPAIALVVTII-----LAKIVSKKNNDISY--DSEEQVGS-DLPAFLPAISGP 227
 + +I+ +I + + LAK +++ +Y D+E QV + +LP FL +I
 Sbjct: 188 AAIPSIIVGSIGCVAFGLFYMKYCLAKSMARGETYATYAFDNEIQVKTKNLPHFLASILPL 247
 Query: 228 LVVICLLALRPLEG-----ITIDPLIALPLGLISILATGYLKETVPFVEYGLSKVVG 280
 L++I + LFG I L+A+ L S L ++ + G S +
 Sbjct: 248 LLLLLIIIALTGSLFGNDFFKKNIIFIALLAVIL--TASWLFQFIPNKIAVFNLGASSSIA 305
 Query: 281 ---VSILLIGTGLSGI IKASNLQFMDIHLLEFLNMPTFILAPLSGIFMGAATASTTSGT 337
 + + G + I+ D+I L P LA L+ M A T S++
 Sbjct: 306 PIFATASAVAFGAVVMIVPGFTFFSDLI--LNIPGNPLISLAVLTS-SMSAITGSSSGAL 362
 Query: 338 TIASQTFATILIKSGVPAVSGAAMIHAGATVLDL 372
 I FA+ + G+ MIH AT+ ++
 Sbjct: 363 GIVMPNFAQYYLDQGL----NPMEIHRVATIASNI 393

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2435-

Example 2160

A DNA sequence (GBSx2277) was identified in *S.agalactiae* <SEQ ID 6677> which encodes the amino acid sequence <SEQ ID 6678>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -3.24    Transmembrane    85 - 101 ( 84 - 101)

----- Final Results -----
10      bacterial membrane --- Certainty=0.2296(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB16041 GB:Z99124 similar to hypothetical proteins [Bacillus subtilis]
    Identities = 176/377 (46%), Positives = 234/377 (61%), Gaps = 2/377 (0%)

Query: 1  MKVVVAIDSLKGSLSLEAGNAIKESINEVISGADVEVHPLADGGEGTVEALTLGMGGTI 60
      MK+++A DS K SLS+LEA AI+ V GAD P+ADGGEGTV++L G I
Sbjct: 1  MKIIIPDSFKESLSALEAAEAIERGFKSVFPGADYRKLPPVADGGEGTVQSLVDATNGRI 60

20  Query: 61 ETIPVKGPLGEKVHASYGIIIPQRLAIEMAAAAGITLIATEERNPLHTTTYGVGEMIKD 120
      V GPLGE V A +G++ + A+IEMAAA+G+ L+ ++RNPL TTT G GE+I
Sbjct: 61 IEQVVIGPLGEPVRAFFGMMGDGRTAVIEMAAASGLHLVPVDKRNPLITTTTGTGELIGA 120

25  Query: 121 AISKGCGRHFIIGIGGSATNDGGAGMLQALGYALLDKDNQEISLGAQGLADLKSISTDKVI 180
      A+ G IIGIGGSATNDGGAGM+QALG LLD EI G L+ L SI +
Sbjct: 121 ALDAGAERLIIGIGGSATNDGGAGMIQALGGRLLDNSGSEIGFGGALSQLASIDVSGLD 180

30  Query: 181 EELKECDFKIIACDVNPLCGAQGCSSIFGPKQGADEDMITKMDTWLSNYATLATSVSEKA 240
      L+ ++AC+V NPL G +G +++FGPQKGA DM+ +D +S++A +A
Sbjct: 181 SRLRNVKLEVACNVNDNPLTGPKGATAVFGPQKGATADMLDVLQNVSHFADMAEKALGST 240

35  Query: 241 DATIEGTGAAGGLGFAFLAFTNATLEPGIDIILSEINIEKAISEADLVVTGEGRLDGQTV 300
      EG GAAGGLG++ L + A L+ GIDI+L ++ E + +ADLV+TGEGR+D QTV
Sbjct: 241 FRDTEGAGAAGGLGWSLLTYLQADLKRGIIDIVLEAVDFESIVQDADLVITGEGRIDSQTV 300

40  Query: 301 MGKAPIGVAKLAKYKGVVAFSGSVTEDAILCNQHGIDAFFPIVRRILSLDEAMSKEVA 360
      GK PIGVAK AK Y V+ +GS++ D+ QHGIDA F IV + L++A
Sbjct: 301 HGKTPIGVAKAAKSYDVPVIGIAGSISRDSNAVYQHGIDALFSIVPGAVPLEDAFEHAAE 360

Query: 361 YKNMKETATQVFRLLNL 377
      Y M+ TA + I L
Sbjct: 361 Y--MERTARDIAASIKL 375

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6679> which encodes the amino acid sequence <SEQ ID 6680>. Analysis of this protein sequence reveals the following:

```

Possible site: 49

50  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.27    Transmembrane    360 - 376 ( 360 - 376)

----- Final Results -----
55      bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

60  >GP:AAA57927 GB:U18997 ORF_f408 [Escherichia coli]
    Identities = 115/345 (33%), Positives = 182/345 (52%), Gaps = 25/345 (7%)

Query: 24  MKILVAIDSFKGSVTSPELNTSVAQALLSVDKQLVIETRAIADGGEGSLVALSQTIVAGRW 83

```

-2436-

MKI++A DS+K S+++ E+ ++ + + + +ADGGEG++ A+ G
 Sbjct: 28 MKIIVIAPDSYKESLSASEVAQAIEKGFREIFPDAAQYVSVPVADGGEGTVEAMIAATQGA 87
 Query: 84 HQVKTIDLLRRPIKVAY--YRHAKQAFIESASIIIGIDKITSNSVTYAQATSYGLGLAVKD 141
 L + ++ K AFIE A+ G++ + + TS G G +
 Sbjct: 88 RHAWVTGPLGEKVNASWGISGDGKTAFIEMAAASGLELVPAEKRDPLVITSRGTGBLILQ 147
 Query: 142 AIQKGATQIEIMLGGTGTSDGGKGFLESINLYDFMT-----GRSYLDTLASPVTLGL 193
 A++ GAT I I +GG+ T+DGG G +++L G L+TL + + + GL
 Sbjct: 148 ALESGATNIIIGIGGSATNDGGAGMVQALGAKLCDANGNEIGFGGSLNLT-NDIDISGL 206
 Query: 194 T-----DVTNPHYGPQGFAAVFGPQKGGSLSQIEETDQIASNFAKKVFCQTTI 241
 DVTNP G G + +FGPQKG S + I E D S++A+ + +
 Sbjct: 207 DPRLKDCVIRVACDVTNPLVGDNGASRIFGPQKGASEAMIVELDNLSHYAEVIKKALHV 266
 Query: 242 DLQTIPTGSGAAGGLGGAIV-LLGGTILTSGFSTRIAELLNLDNSLQSCDLVITGEGCLDTQS 300
 D++ +PG+GAAGG+G A++ LG L SG + LNL+ + C LVITGEG +D+QS
 Sbjct: 267 DVKDVPGAGAAGGMGAALMAFLGAEKSGIEIVTTALNLEBHIDCTLVITGEGRIDSQS 326
 Query: 301 QSGKVPVAIARMAKKYQVPTIALCGSVKIETGLAEDFL-AVFSI 344
 GKVP+ +A +AKKY P I + GS+ + G+ + + AVFS+
 Sbjct: 327 IHGKVPIGVANVAKKYHKPVIGIAGSLTDDVGVVHQHGIDAVFSV 371

An alignment of the GAS and GBS proteins is shown below.

Identities = 128/379 (33%), Positives = 194/379 (50%), Gaps = 23/379 (6%)
 Query: 1 MKVVVAIDSLKGLSSLEAGNAIKESINEVISGADVEVHPLADGGEGTVEALTGMGGTI 60
 MK++VAIDS KGS++S E ++ +++ V +E +ADGGEG++ AL+ + G
 Sbjct: 24 MKILVAIDSFKGSVTSPELNTSVAQALLSVDKQLVIEITRAIDGGEGSLVALSQTAVGRW 83
 Query: 61 ETIPVKGPLGEKVHASYGIIIPQRLAIIEMAAAGITLIATEERNPLHTTTYGVGEMIKD 120
 + L + +Y + A IE A+ GI I + T+YG+G +KD
 Sbjct: 84 HQVKTIDLLRRPIKVAY--YRHAKQAFIESASIIIGIDKITSNSVTYAQATSYGLGLAVKD 141
 Query: 121 AISKGRHFIIIGIGGSATNDGGAGMLQALGYALLDKDNQIEISLGAQGLADLKSISTDKVI 180
 AI KG I +GG+ T+DGG G L++L Y + G + L ++++ +
 Sbjct: 142 AIQKGATQIEIMLGGTGTSDGGKGFLESINLYDFMT-----GRSYLDTLASPVTL 190
 Query: 181 EELKECDFKIACDVTNPLCGAQGCSSIFGPQKGADEDMITKMDTWLSNYATLATSVEKA 240
 L DVTNP G QG +++FGPQKG I + D SN+A +
 Sbjct: 191 LGLT-----DVTNPHYGPQGFAAVFGPQKGGSLSQIEETDQIASNFAKKVFCQTTID 242
 Query: 241 DATIEGTGAAGGLGFAFLAFTNATLEPGIDIILSEINIEKAISEADLVVTGEGRLDGQTV 300
 TI G+GAAGGLG A + TL G I +N++ ++ DLV+TGEGLD Q+
 Sbjct: 243 LQTIPGSGAAGGLGGA-IVLLGGTILTSGFSTRIAELLNLDNSLQSCDLVITGEGCLDTQSQ 301
 Query: 301 MGKAPIGVAKLAKKYGKKVAFSGSVTEDAILCNQHGDIAFFPIVRRRLISLDEAMSKEVA 360
 GK P+ +A++AKKY +A GSV + L + + A F I ++ ISL+ A+ K
 Sbjct: 302 SGKVPVAIARMAKKYQVPTIALCGSVKIETGLAEDFL-AVFSIQQPISLEAAIDKTTT 360
 Query: 361 YKNMKETATQVFRLINLYN 379
 N+K A + LI +N
 Sbjct: 361 LSNIKILAAANMLLIAQFN 379

SEQ ID 6678 (GBS409) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 7; MW 45.4kDa).

GBS409-His was purified as shown in Figure 214, lane 6.

GBS409d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 3 & 4; MW 35kDa) and in Figure 188 (lane 12; MW 35kDa). Purified protein is shown in Figure 240, lanes 9-10.

-2437-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2161

A DNA sequence (GBSx2278) was identified in *S.agalactiae* <SEQ ID 6681> which encodes the amino acid sequence <SEQ ID 6682>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1886(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC21771 GB:U32695 conserved hypothetical protein [Haemophilus influenzae Rd]
Identities = 97/383 (25%), Positives = 175/383 (45%), Gaps = 52/383 (13%)

Query: 1 MKLRKQLAQQIVTSIKDVCQQDINFINTKGIIFASTNPKRVGEPHEIGLKVAQTGQMIEV 60
M+L K A++IV + +N ++ G+I AS N R+ + H + + +++E+
20 Sbjct: 1 MQLDKYTAKKIVKRAMKIIHHSVNVMDHDGVI IASGNSTRINQRHTGAVLALRENRVVEI 60

Query: 61 TD---QESYFGTQAGINIPFYNCCELLATIGISGNPNQVGKYALLAQKMTRLILKEHE-L 116
Q+ F Q GIN+P +Y + + +GISG P QV +YA L + LI+++ L
25 Sbjct: 61 DQALAQKWNFEAQPGINLPIHYLGKNIGVVGISGEPTQVKQYAEVLKMTAELIVEQQALL 120

Query: 117 DYLDGFRKNEASIVLHHLVEGRELDYYLNOFLNOYHLSEKTDYRLLTFEINSQKQKLLL 176
+ + R+ + +L L+ LN + ++ + +F++N + +L+
25 Sbjct: 121 EQESWHRRYKEEFILQ-----LLHCNLNWKEMEQA--KFFSFDLNKSRVVVLI 167

Query: 177 S-----QSEMSLLNFFDK-----LDTAIYTFNYPNQYWLILLSDHMFDDYYYPNI 219
+ +L+N+ ++ LD + + N +LS M
30 Sbjct: 168 KLLNPALDNLQNLINYLEQSEFAQDVAILSLDQVVVLKTWQNS--TVLSAQM-----KT 219

Query: 220 LSKFECEKGLYKVGIGQKSSLSLLKR---SYETSILALK-ALKGQOK--VNLVDDLDLEL 273
L + K YK+ +G +L L ++ S++++ L LK + + + D+ L +
35 Sbjct: 220 LLPADYSKQDYKIAVGACLNLPFLFEQLPLSFQSAQSTLSYGLKHHPRKGIYVDFEHLRPV 279

Query: 274 LLTSIDSNIKQVVLNKAIVNL-SENDKIL---LNSYFKHNLSLKECSQELFIHKNTVQYR 329
LL + + + L K L L SE + IL L YF N L +++LF+H NT++YR
40 Sbjct: 280 LLAGLSHWSQGNELIKPLSPLFSEENAILYKTLQYFSLNCDLYLTAEKLFVHPNTLRYR 339

Query: 330 LNKIYESTQLNPRNFKDATLLYL 352
LNKI + T L D LYL
45 Sbjct: 340 LNKIEQITGLFFNKIDDKLTLYL 362

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2162

50 A DNA sequence (GBSx2279) was identified in *S.agalactiae* <SEQ ID 6683> which encodes the amino acid sequence <SEQ ID 6684>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.0290(Affirmative) < succ>

-2438-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF89979 GB:AF206272 beta-glucosidase [Streptococcus mutans]
 Identities = 334/475 (70%), Positives = 392/475 (82%), Gaps = 8/475 (1%)

Query: 4 FPKHFLWGGAVAANQVEGAFRTDGKGLSVQDVLNPGGLGD-----FTAKTPDNLKLE 56
 FP++FLWGGGA AANQ EGA+ DGKGLSVQDV P GG+ T KPT DNLKL

10 Sbjct: 6 FPFNFLWGGATAANQFEGAYNQDGKGLSVQDVTPKGGVAQSGSSSPLITEKPTEDNLKLV 65

Query: 57 AIDFYHNYKNDIKLFAEMGFKVFRTSIAWSRIFPNGDDSAPEAGLQFYDNLDFDELLKYN 116
 IDFY+ YK DI LFAEMGFKVFR SIAW+RIFPNGDD PNEAGL FYD +FDEL KY+

15 Sbjct: 66 GIDFYNRYKEDIALFAEMGFKVFRLSIAWTRIFPNGDDLEPNEAGLAFYDKVDFDELA KYD 125

Query: 117 IEPLVTLSHYETPLHLAKTYNGWADRRLIAFFEKFAQTMERYKDKVKYWLTFNEVNSIL 176
 IEPLVTLSHYETPLHLA+ YNGWA+R LIAF+E++A+TV RYKDKVKYWLTFNEVNS+L

Sbjct: 126 IEPLVTLSHYETPLHLARKYNGWANRELIAFYERYARTVFTRYKDKVKYWLTFNEVNSVL 185

20 Query: 177 HMPFTSGAIMTDKSQLSPQELQAIHHELVASARVTKLGRSINPNFKIGCMILAMPAYPM 236
 H PF SG I+TD QLS Q+LYQA+HHELV SA TK+G INP+FKIGCM+LAMPAYPM

Sbjct: 186 HAPFMSGGIITDPEQLSKQDLYQAVHHELVVSALATKVGHEINPDFKIGCMVLAMPAYPM 245

25 Query: 237 TSDPRDVLAAARQFEQHNLFLSDIHVRGKYPTYIQSYFKNNGIKIKFEEGDEEVLAQNTVD 296
 T+DP D LA R+FE N LFSH+H RGKYP YI+ YFK+N I IK EGD+E++ +NTVD

Sbjct: 246 TADPLDQLAVREFENQNYLFSDLHARGKYPNYIKRYFKDNNIDIKMGEQDKELMLENTVD 305

Query: 297 FLSFSYYMSVTOAYDFENYQSGQGNILGGLTNPHLTTSEWGQIDPIGLRLVLNQYYERY 356
 F+SFSYYMSV A++ E+Y SG+GN+LGGL+NP+L SEWGQIDP+GLRLVLN Y+RY

30 Sbjct: 306 FISFSYYMSVAAAHNPEDYNSGRGNVLGGLSNPYLQASEWGQIDPVGLRLVLNDSYDRY 365

Query: 357 QIPLFIVENGLGAKDQLIETLDGDTVEDDYRIDYMNQHLVQVAKAIEDGVEIMGYTSWG 416
 Q+PLFIVENGLGAKD L++ DG TVEDDYRIDY+ +HL+QV +A++DGV+++GYT+WG

35 Sbjct: 366 QIPLFIVENGLGAKDVLVQGPDP-TVEDDYRIDYLQKHLMQVGEALQDGVDLLGYTTWG 424

Query: 417 CIDCVSMSTAQLSKRYGLIYVDRNDDGTGSLQRYKKKSFQWYQKVIKTNGQSLFE 471
 ID VS ST +LSKRYG IYV NDDG+GSL RYKKKSF WY+KVI+TNG SL+E

Sbjct: 425 PIDLVSESTVELSKRYGFIYVACNDDGSGSLARYKKKSPAWYKKVIETNGASLYE 479

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5287> which encodes the amino acid sequence <SEQ ID 5288>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0763(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 390/469 (83%), Positives = 423/469 (90%)

Query: 1 MTVPFKHFLWGGAVAANQVEGAFRTDGKGLSVQDVLNPGGLGDF TAKTPDNLKLEAIDF 60
 M +FPK FLWGGAVAANQVEGAF D KGLSVQDVLNPGGLG++T PT DNL LEAIDF

55 Sbjct: 1 MGIFPKDFLWGGAVAANQVEGA FEADAKGLSVQDVLNPGGLGEWTDSP TSDNL TLEAIDF 60

Query: 61 YHNYKNDIKLFAEMGFKVFRTSIAWSRIFPNGDDSAPEAGLQFYDNLDFDELLKYNIEPL 120
 YH YK DI LFAEMGFKVFRTSIAWSRIFPNGDD PNEAGLQFYD+LFEDELL Y IEPL

60 Sbjct: 61 YHRYKEDIALFAEMGFKVFRTSIAWSRIFPNGDDQ PNEAGLQFYDDLFDELLNYGIEPL 120

Query: 121 VTLSHYETPLHLAKTYNGWADRRLIAFFEKFAQTMERYKDKVKYWLTFNEVNSILHMPF 180
 VTLSHYETPLHLAK YNGW DRRLI FFE+FAQTMERYKDKVKYWLTFNEVNSILHMPF

Sbjct: 121 VTLSHYETPLHLAKAYNGWTDRLIGFFERFAQTMERYKDKVKYWLTFNEVNSILHMPF 180

-2439-

Query: 181 TSGAIMTDKSQLSPQELYQAIHHELVASARVTKLGRSINPNFKIGCMILAMPAYPMTSDP 240
 TSG IMT+K +LS Q+LYQAIHHELVASA VTKL INP+ K+GCMILAMPAYPMTSDP
 Sbjct: 181 TSGGIMTEKEKLSLQDLYQAIHHELVASASVTKLAHEINPDVKVGCMLAMPAYPMTSDP 240

5 Query: 241 RDVLAARQFEQHNLFSDIHVRGKYPTTYIQSYFKNNGIKIKFEEGDDEEVLQONTVDFLSF 300
 RD+LAA FE NLLFSDIHVRGKYP+YI+SYFK NGI+I FE+GD+E+LA++TVDFLSF
 Sbjct: 241 RDILAAHAFENLNLFSDIHVRGKYPYSYKSYFKENGIEIVFEDGDKELLAETVDFLSF 300

10 Query: 301 SYYSVVTQAYDFENYQSGQGNILGGLTNPHLTSEWGWQIDPIGLRLVLNQYYERYQIPL 360
 SYYSVVTQA++ E Y SGQGNILGGL+NP+L +SEWGWQIDPIGLRLVLNQYY+RYQIPL
 Sbjct: 301 SYYSVVTQAHNPEAYTSGQGNILGGLSNPYLESSEWGWQIDPIGLRLVLNQYYDRYQIPL 360

15 Query: 361 FIVENGLGAKDQLIETLDGDYTVEDDYRIDYMNQHLVQVAKAIEDGVEIMGYTSWGCIDC 420
 FIVENGLGAKDQL++T DG TV DDYRIDYM+QHLVQVAKAIEDGVE+MGYTSWGCIDC
 Sbjct: 361 FIVENGLGAKDQLVQTADGSMTHDDYRIDYMSQHLVQVAKAIEDGVEVMGYTSWGCIDC 420

20 Query: 421 VSMSTAQLSKRYGLIYVDRNDDGTGSLQRYKKKSGWYQKVIKTNGQSL 469
 VSMSTAQLSKRYG IYVDRNDDGTG L RYKKKSF WY++VI+TNG+ L
 Sbjct: 421 VSMSTAQLSKRYGFIYVDRNDDGTQLTRYKKKSFWDYRQVIQTNGRYL 469

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2163

A DNA sequence (GBSx2280) was identified in *S. agalactiae* <SEQ ID 6685> which encodes the amino acid sequence <SEQ ID 6686>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.40	Transmembrane	247 - 263 (241 - 273)
INTEGRAL	Likelihood = -8.55	Transmembrane	429 - 445 (424 - 450)
INTEGRAL	Likelihood = -4.88	Transmembrane	285 - 301 (280 - 303)
INTEGRAL	Likelihood = -3.82	Transmembrane	207 - 223 (205 - 225)
INTEGRAL	Likelihood = -3.40	Transmembrane	113 - 129 (112 - 139)
INTEGRAL	Likelihood = -1.97	Transmembrane	309 - 325 (305 - 328)
INTEGRAL	Likelihood = -1.59	Transmembrane	395 - 411 (395 - 411)
INTEGRAL	Likelihood = -1.49	Transmembrane	174 - 190 (173 - 193)

----- Final Results -----

bacterial membrane	---	Certainty=0.5161(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA84286 GB:Z34526 beta-glucoside permease [Bacillus subtilis]
 Identities = 225/594 (37%), Positives = 351/594 (58%), Gaps = 11/594 (1%)

45 Query: 4 YQETAKAILAAVGGKNIQHVTCHVTRLRLVLNDDEIVNDQVIKTIPNVIGVMRKNDQYQ 63
 Y + +K IL VGGE+N+Q V HC+TRLR L ++ + ++ +P V+G +Q+Q
 Sbjct: 3 YDKLSKDILQLVGGEENVQRVIHCMTRLRFNLHDNAKADRSQLEQLPGVMGTNISGEQFQ 62

50 Query: 64 IILGNDVNNYNAFLALGHFENTTREFSSQKKSSILEKLIETIAGVITPLIPALLGGGML 123
 II+GNDV Y A + + + SS +K ++L + + I+GV TP++PA+ G GM+
 Sbjct: 63 IIGNDVPKVYQAIVRHSNLSDEKSAGSSSQKKNVLSAVFDVISGVFTPLIPAAGAGMI 122

55 Query: 124 KVIGILLPMLGIASSSSQTVAFINFFGDAAYYFMPIMIAYSAASRFKVTPLAATVGGIL 183
 K + L G + SQ + GD A+YF+P++++A SAA +F P +AA + +
 Sbjct: 123 KGLVALAVTFGWMAEKSQVHVILTAVGDGAFYFLPLLAMSAARKFGSNPYVAAAIAAAI 182

60 Query: 184 LHPAFVTMVAEGKPLSLFGAPVTLASYGSSVIPILIMVFLMQYIERWINKIVPSVMKSFL 243
 LHP ++ GKP+S G PVT A+Y S+VIPIL+ +++ Y+E+WI++ + +K +
 Sbjct: 183 LHPDLTALLGAGKPISFIGLPVTAATYSSTVIPILLSIWIASYVEKWIDRFTHASLKLIV 242

Query: 244 QPTLIILISGFLALVVVGPLGVIIGKGLSSAMLSIYHVAPWLALSILGAIMPLVVMGMH 303

-2440-

PT +LI L L+ VGPLG I+G+ LSS + ++ A +A+ +L L++MTGMH
 Sbjct: 243 VPTFTLLIVVPLTLITVGPLGAILGEYLSSGVNLYFDHAGLVAMILLAGTFSLIIMTGMH 302

5 Query: 304 WAFAPIFLAASVATPDVLIILPAMLASNLAQGAASLAVAVKAKQKQTRQVAFAGLSALLA 363
 +AF PI + +LPAM +N+ Q AS AV ++++ K+ + +A ++AL+
 Sbjct: 303 YAFVPIMINNIAQNQHLYLLPAMFLANMGQAGASFAVFLRSRKKFKSLALTTITAILM- 361

10 Query: 364 GITEPALYGVTLKFKKPLYAAMISGGLVGAYIGLVNIASTYTFVVPSTIIGLPQYINPQGGN 423
 GITEPA+YGV ++ KKP AA+I G GA+ G+ +ASY +V GLP I G
 Sbjct: 362 GITEPAMYGVNMLKKPFAAALIGGAAGGAFYGMTGVASY--IVGGNAGLPS-IPVFIGP 418

15 Query: 424 NFSNAVIAAIIATIIITFIITWFLGIDEGENEKSSINAQETHIRSGLSKSKETLYSPMVGN 483
 F A+I + + LG ++ ++ S Q H S +E ++SP+ G
 Sbjct: 419 TFIYAMIGLVIAFAAETAAYLLGFEDVPSDGSQ---QPAVHEGS---REIIHSPIKGE 471

20 Query: 484 VLPLSKVPDETFFSKLLGEGLAITPSVGEVYAPFDGEIISLFPPTKHAIALKDDKGVEVLI 543
 V LS+V D FS+ ++G+G AI P GEV +P G + ++F TKHAI + D+G E+LI
 Sbjct: 472 VKALSEVKDGVFSAGVMGKGFAIEPEEGEVVSPVRGVTITPKTKHAIGITSDQGAEILI 531

25 Query: 544 HIGIDTVELNGEGFEQLVKVGDVFKRGQLLLRMDIDFISCKGYSLISPVVVTNS 597
 HIG+DTV+L G+ F +K GD V G L+ D++ I + GY +I+PV+VTN+
 Sbjct: 532 HIGLDTVLEKQWFTAHKEGDKVAPGDPLVSPDLEQIKAAGYDVITPVIVTNT 585

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2883> which encodes the amino acid
 25 sequence <SEQ ID 2884>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.40	Transmembrane	246 - 262 (240 - 271)
INTEGRAL	Likelihood = -6.26	Transmembrane	284 - 300 (279 - 304)
INTEGRAL	Likelihood = -4.14	Transmembrane	173 - 189 (172 - 194)
INTEGRAL	Likelihood = -3.24	Transmembrane	112 - 128 (111 - 137)
INTEGRAL	Likelihood = -2.39	Transmembrane	428 - 444 (425 - 445)
INTEGRAL	Likelihood = -2.13	Transmembrane	383 - 399 (380 - 401)
INTEGRAL	Likelihood = -1.97	Transmembrane	308 - 324 (304 - 327)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 508/619 (82%), Positives = 561/619 (90%), Gaps = 1/619 (0%)

45 Query: 4 YQETAKAILAAVGGKNIQHVTCHVTRLRLVLDNDEIVNDQVIKTIPNVIGVMRKNDQYQ 63
 YQETAKAILAAVGG+ NIQ VTHCVTRLRLVL NDE V DQ +K I NVIGVMRKN QYQ
 Sbjct: 3 YQETAKAILAAVGGKTNIQRVTHCVTRLRLVLKNDEKVKDQVKAISNVIGVMRKNQYQ 62

50 Query: 64 IILGNDVNNYNAFLALGHFENTTREFSSQKKSSILEKLIETIAGVITPLIPALLGGGML 123
 IILGNDVNNYY AFL+LGHF+N + SS+ K SILE+LIETIAGVITPLIPALLGGGML
 Sbjct: 63 IILGNDVNNYYQAFSLSLGHFDNQDEHSSKAKGSILERLIETIAGVITPLIPALLGGGML 122

55 Query: 124 KVIGILLPMLGIASSSSQTVAFINFFGDAAYYFMPIMIAYSASARFKVTPVLAATVGGIL 183
 KV+GILLPMLG+AS+ SQTVAFINFFGDAAYYFMP+MIAYSAA+RFKVTPLAAT+ GIL
 Sbjct: 123 KVVIGILLPMLGLASADSQTVAFINFFGDAAYYFMPVMIAYSAAARFKVTPVLAATIAGIL 182

60 Query: 184 LHPAFVTMVAEGKPLSLFGAPVTLASYGSSVIPILIMVFLMQYIERWINKIVPSVMKSFL 243
 LHPAFV MVAEGKPL+LFGAPVT ASYGSSVIPIL+MV+LMQYIE+W+N++VPSVMKSFL
 Sbjct: 183 LHPAFVAMVAEGKPLTLFGAPVTPASYGSSVIPILMMVYLMQYIEKWNRLVPSVMKSFL 242

65 Query: 244 QPTLIILISGFLALVVVGPLGVIIGKGLSSAMLSIYHVAPWLALSILGAIMPLVVMTGMH 303
 QPTLIILISGFLALVVVGPLGVIIG+GLS+ ML+IYHVAPWLAL+ILGAIMPLVVMTGMH
 Sbjct: 243 QPTLIILISGFLALVVVGPLGVIIGQLSNTMLAIYHVAPWLALAILGAIMPLVVMTGMH 302

Query: 304 WAFAPIFLAASVATPDVLIILPAMLASNLAQGAASLAVAVKAKQKQTRQVAFAGLSALLA 363
 WAFAPIFLAASVATPDVLIILPAMLASNLAQGAASLAVA K KQKQTRQVA AAG+SALLA

-2441-

Sbjct: 303 WAFAPIFLAASVATPDVILPAMLASNLAQGAASLAVAFKTKQKQTRQVALAAGISALLA 362

Query: 364 GITEPALYGVTLKFKKPLYAAMISGGLVGAYIGLVNIASTYTFVVPSTIIGLPQYINPQGGN 423
GITEPALYGVTLKFKKPLYAAMISGGLVGA+IG VNIASTYTFVVPSTIIGLPQYINP GG

5 Sbjct: 363 GITEPALYGVTLKFKKPLYAAMISGGLVGAFIGFVNIASTYTFVVPSTIIGLPQYINPSCGA 422

Query: 424 NFSNAVIAAIAIITLFTIITWFLGIDEGENEKSSINAQETHIRSGLSKKTLYSPMVGN 483
NF+NA+IA ATI+L F +TWF+GIDE E+ K A + + ++SGLS K+TLY+PM G

10 Sbjct: 423 NFTNALIAGTATIVLAFSLTWFMGIDE-ESPKQVSVAADMSQVKSGLSTKQTLTAPMTGE 481

Query: 484 VLPLSKVPDETFSKLLGEGLAITPSVGEVYAPFDGEIISLFPTKHAIALKDDKGVEVLI 543
+L LS+VPDETFSKLLGEG AI PS GEVYAPFDGE+I+ FPTKHA+ALK+ +GVEVLI

Sbjct: 482 MLFLSEVPDETFSKLLGEGFAILPSEGEVYAPFDGEVITFFPTKHAVALKNTRGVEVLI 541

15 Query: 544 HIGIDTVELNGEGFEQLVKVGDFVKRGQLLLRMDIDFISKGYSLISPVVVTNSIDQLEI 603
H+GIDTVEL G+GFEQLV VGD VKRGQ LL+MDIDFI+SKGYSLISPVVVTNS +QLEI

Sbjct: 542 HVGIDTVELKGQFEQLVSVGVDVVKRGQALLKMDIDFITSKGYSLISPVVVTNSAEQLEI 601

Query: 604 IVKDAETMVINEDDLVL 622
I++D + MVT ED LLVIL

20 Sbjct: 602 IIQDDKMMVTKEDALLVL 620

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2164

A DNA sequence (GBSx2281) was identified in *S.agalactiae* <SEQ ID 6687> which encodes the amino acid sequence <SEQ ID 6688>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1148(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15944 GB:Z99124 transcriptional antiterminator (BglG family)
[Bacillus subtilis]
Identities = 118/275 (42%), Positives = 183/275 (65%)

40 Query: 1 MIKRVLNHNNAVISVTHQGLDVLMLGKGIAPKKRIGDRINSDAIEKSFVLKNSDNMNRFT 60
M I +V+N+N + V QG ++++MG+G+AF+K+ GD ++ IEK F L N D +F
Sbjct: 1 MKIAKVINNNVISVNEQGKELVVMGRGLAFQKKSGDDVDEARIEKVFTLDNKNVSEKFK 60

45 Query: 61 ELFITVPEEVVACSERIINLGKIKLGNLDEILYINLTDHIHSAIERHEQGMVIONPLRL 120
L +P E + SE II+ K++LGK L++ +Y++LTDHI+ AI+R+++G+ I+N L
Sbjct: 61 TLLYDIPIECMEVSEETIHYAKLQLGKKLNDSTYVSLTDHINFATQRNQKGLDIKNALLW 120

50 Query: 121 EIQRYPDEYSIGMKALELIKDELGICLTIDESAFIAMHFVNAGLDNPFNEAHKITEIVS 180
E +R Y DE++IG +AL ++K++ G+ L DE+ FIA+H VNA L+ IT+++
Sbjct: 121 ETKRLYKDEFAIGKEALVMVKNKTGVSLPEDEAGFIALHIVNAELNEEMPNIINITKVMQ 180

55 Query: 181 YIEQVKIDFRTELDESSIDYYRFMTHTKLFQARVLSGMKYEDDDADLLLVVKKKYPREY 240
I VK F+ E +E S+ YYRF+TH K FAQR+ +G E D LL VK+KY R Y
Sbjct: 181 EILSVKYHFKIEFNEESLHYRFVTHLKFFAQRLLFNTHMESQDDFLLDTVKEKYHRAY 240

Query: 241 KCVKEIGNNMAIQYQYQLNSSELLYLTVHVKRLVK 275
+C K+I + +Y+++L S ELLYLT+H++R+VK

60 Sbjct: 241 ECTKKIQTYIEREYEHKLTSEDELLYLTIHIERVVK 275

-2442-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6689> which encodes the amino acid sequence <SEQ ID 6690>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0680(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 220/279 (78%), Positives = 246/279 (87%)

Query: 1 MIIKRVLNHNAVISVTHQGLDVLMLMGKGI AFKKRIGDRINSDAIEKSFVLKNSDNMNRFT 60
 M+IKRVLNHNA IS HQGLD+LLMGKGI F K++GD I +AIE SFVLKNSDNMNRFT
 Sbjct: 1 MLIKRVLNHNAISTNHQGLDILLMGKGITFGKKVGDSEIENALTSFVLKNSDNMNRFT 60

Query: 61 ELFITVPEEVVACSERIINLGKIKLGKLNDEILYINLTDHIHSAIERHEQGMVIONPLRL 120
 ELFITVPEEVVACSERIINLGKIKLGK LNDEILYINLTDHIHSAIERHEQGM+I NPLR
 Sbjct: 61 ELFITVPEEVVACSERIINLGKIKLGKLNDEILYINLTDHIHSAIERHEQGMVIONPLRL 120

Query: 121 EIQRYPDEYSIGMKALELIKDELGICLTIDESAFIAMHFVNAGLDNPFNEAHKITEIVS 180
 EIQRYPDEYS+G+KALELI+ LG+ L IDE+AFIAMHFVNA LD PF E H++TEIVS
 Sbjct: 121 EIQRYPDEYSLGVKALELIERNLGVTLAIDEAAFIAMHFVNASLDTPFKEPHRLTEIVS 180

Query: 181 YIEQKVKIDFRTELDESSIDYRFMTHTKLFAQRVLSGMKYEDDDADLLLVVKKYPREY 240
 YIEQK+K DF+TELD++SIDYRFMTHT KLFAQRVLS M Y+DDDA+LLLVVK KYP+EY
 Sbjct: 181 YIEQKIKTDFKTELDDESSIDYRFMTHTKLFAQRVLSQMSYDDDDABLLLVVKKYPKEY 240

Query: 241 KCVKEIGNNMAIQYQYQLNSSELLYLTVHVKRLVKNLKE 279
 +CV +I + +Y Y LNSSELLYLTVHVKRLVK+LKE
 Sbjct: 241 RCVLDISEBIKKRYNYHLNSSELLYLTVHVKRLVKHLKE 279

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2165

A DNA sequence (GBSx2282) was identified in *S.agalactiae* <SEQ ID 6691> which encodes the amino acid sequence <SEQ ID 6692>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1104(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9335> which encodes amino acid sequence <SEQ ID 9336> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6693> which encodes the amino acid sequence <SEQ ID 6694>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3314(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2443-

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/178 (80%), Positives = 161/178 (90%)

```

5   Query: 1   MTLHHDKHHATYVANANAALKHPEIGEDLEALLADVSQIPEDIRQAVINNGGGHLNHAL 60
      MTLHHDKHHATYVAN NAALEKHPEIGE+LE LLADV++IPEDIRQ +INNGGGHLNHAL
      Sbjct: 24 MTLHHDKHHATYVANTNAALEKHPEIGENLEELLADVTKIPEDIRQTLINNGGGHLNHAL 83

10  Query: 61  FWELMSPEETQISQELSEDINATFGSFEDFKAAFTAAATGRFGSGWAWLVVNAEGKLEVL 120
      FWEL+SPE+ ++ ++++ I+ FGSF+ FK FTAAATGRFGSGWAWLVVN EG+LE+
      Sbjct: 84 FWELLSPEKQDVTDPDVAQAIDDAFGSFDAFKEQFTAAATGRFGSGWAWLVVNKEGQLEIT 143

      Query: 121 STANQDTPIMEGKKPILGLDVWEHAYYLNRYNRVNPNIKAFFEINWNKVNELYQAAK 178
      STANQDTP I EGKKPIL LDVWEHAYYLNRYNRVNPNIKAFFEI+NW KV+ELYQAAK
15  Sbjct: 144 STANQDTPISEGKKPILALDVWEHAYYLNRYNRVNPNIKAFFEIVNKKVSELYQAAK 201

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2166

20 A DNA sequence (GBSx2283) was identified in *S.agalactiae* <SEQ ID 6695> which encodes the amino acid sequence <SEQ ID 6696>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

```

```

25  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3331(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2167

35 A DNA sequence (GBSx2284) was identified in *S.agalactiae* <SEQ ID 6697> which encodes the amino acid sequence <SEQ ID 6698>. This protein is predicted to be DNA polymerase III delta subunit. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

```

```

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0511(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9743> which encodes amino acid sequence <SEQ ID 9744> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6699> which encodes the amino acid sequence <SEQ ID 6700>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 43

```

-2444-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.22 Transmembrane 250 - 266 (249 - 266)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 222/340 (65%), Positives = 282/340 (82%)

Query: 1 MIAIEEIGRITPDNLGLVTVLAGEDLGGYAQMKEKLFQVIGFNKDDLAYSYPDLSEEDYQ 60
 MIAIE+I +++ +NLGL+T++ G+D+GQY+Q+K +L + I F+KDDLAYSYPD+SE YQ
 Sbjct: 1 MIAIEKIEKLSKENLGLITLVTGDDIGQYSQLKSRLMEQIAFDKDDLAYSYPDMSEAAAYQ 60

15 Query: 61 NAELEDLESLPFLSDYKVVIFDQFQDITTDKKTLYLDEQAMKRFAYLQNPVDTTTRLVICAP 120
 +AE+DL SLPF ++ KVVIFD DITT+KK++L E+ +K FEAYL+NP++TTRL+I AP
 Sbjct: 61 DAEMDLVSLPFFAEQKVVIFDHLDDITTNKKSFLKEKDLKAFEAYLENPLETTRLIIFAP 120

20 Query: 121 GKLDGKRRLVKLLKRDARVLEANTLKESDLKTYFQKYAHQEGLVFEAGVFDELLIKSNYD 180
 GKLD KRLVKLLKRDA VLEAN LKE++L+TYFQKY+HQ GL FE+G FD+LL+KSN D
 Sbjct: 121 GKLDKRRRLVKLLKRDALVLEANPLKEAELRTYFQKYSHQLGLGFESGAFDQLLKSNDD 180

25 Query: 181 FSDTLTNIAFLKSYKTDGHISNSDVREAI PKSLQDNIFDLTQDVLGRIDLARDLVRDLR 240
 FS + N+AFLK+YK G+IS D+ +AIPKSLQDNIFDLT+ VL G+ID ARDL+ DLR
 Sbjct: 181 FSQIMKNMAFLKAYKKTGNISLTDIEQAIPKSLQDNIFDLTRLVLGGKIDAARDLIHDLR 240

30 Query: 241 LOGEDEIKLIAIMLGQFRMFLQVKILASKGKSESQIVSELSHYIGRKINPYQVKFAVRDS 300
 L GED+IKLIAIMLGQFR+FLQ+ ILA K+E Q+V LS +GR++NPYQVK+A++DS
 Sbjct: 241 LSGEDDIKLIAMLGQFRLFLQLTILARDVKNEQQLVISLSDILGRRVNPYQVKYALKDS 300

35 Query: 301 RNLPLAFLKEAIRILIEDYAIKRGTYDKDYLFDLALLKI 340
 R L LAFL A++ LIETDY IK G Y+K YL D+ALLKI
 Sbjct: 301 RTLSLAFLTGAVKTLIEDYQIKTGLYEKSYLVDIALLLKI 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2168

40 A DNA sequence (GBSx2285) was identified in *Sagalactiae* <SEQ ID 6701> which encodes the amino acid sequence <SEQ ID 6702>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3071(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2169

55 A DNA sequence (GBSx2286) was identified in *Sagalactiae* <SEQ ID 6703> which encodes the amino acid sequence <SEQ ID 6704>. This protein is predicted to be esterase. Analysis of this protein sequence reveals the following:

-2445-

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -0.32 Transmembrane 175 - 191 (175 - 191)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1128(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB17013 GB:L38252 esterase [Acinetobacter lwoffii]

Identities = 63/218 (28%), Positives = 107/218 (48%), Gaps = 3/218 (1%)

Query: 105 KVIFYVHGGSYIHQASELQYIFVNKLAKKLDKVVFPYIPKAPTNYSDAIPKIKKLYQN 164

15 ++IF++HGG++ + + LA + +V+ YP AP + Y +AI I +YQ

Sbjct: 73 QLIFHIHGGAFFLGLSLNTHRALMTDLAARTQMQUIHVDYPLAPEHPYPEAIDAIFDVYQA 132

Query: 165 TLASVTSPKQIILVGESAGGGLALGLADNLVTEHIKQPKBIILISPWLDIATNNPKIEKV 224

20 L PK II+ G+S G LAL L L + P +IL+SP+LD+ + +

Sbjct: 133 LLVQGIKPKDIIISGDSCGANLALALCLRLKQQPELMPSSGLILMSPYLDLTLTSESIRFN 192

Query: 225 QKKDPLLKAWQLQQVAPYWANGKKNFKNQVSPLYSSQFNKMAPISFFIGTHDIFYPDNQ 284

25 QK D LL LQ ++ +P+VSPL+ + + P +G+ +I D++

Sbjct: 193 QKHDALLSIEALQAGIKHYLTDDIQPGDERVSPFLF-DDLDGLPPTLVQVGSKEILLDDSK 251

Query: 285 LLHQKLAKENIKHHYIVGQKMNHVYPVLP--IPEAETA 320

25 +K + ++K H+ + M H + + PEA+ A

Sbjct: 252 RFREKAEQADVKVHFVKLYTGMWHNFMFNAWFPEAKQA 289

30 There is also homology to SEQ ID 3498.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2170

A DNA sequence (GBSx2287) was identified in *S.agalactiae* <SEQ ID 6705> which encodes the amino acid sequence <SEQ ID 6706>. This protein is predicted to be purine nucleotide synthesis repressor. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2970(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16124 GB:Z99124 similar to transcriptional regulator (LacI

family) [Bacillus subtilis]

Identities = 111/300 (37%), Positives = 175/300 (58%), Gaps = 4/300 (1%)

50 Query: 1 MTSISDIACKAGVAKSTVSRVINHHPHVSDETRQKVMALITELDYIPNQLARDLSRGKTQ 60

M +I +IA+ A V+ STVSRV+NHHP+VS+E R+ V ++ ELDY PN+ A DL RGKT

Sbjct: 1 MANIKEIARLANVSVSTVSRVLNHHHPVSEEKRLVHVQVMKELDYTPNRTAIDLIRGKTH 60

Query: 61 KIGVVIPHTRHPYFTQLINGLLDAAKTIDYQLVMMPSDYNOEELS YLKQLKMEAIDALI 120

55 +GV++P++ HP F +++NG+ AA +Y ++P++YN ++E+ YL+ L+ + ID LI

Sbjct: 61 TVGVILPYSDHPCFDKIVNGITKAAFQHEYATTLPTNYPNDIEIKYLELLRTKKIDGLI 120

Query: 121 FTSRAISLDIIETIYAKYGRIVVCEKLQFYNHLSSAYLDRYSSFLEAFSDMKLRGLEHLVL 180

TSRA D I Y +YG ++ CE + + + A+ DR +++ E+F +K RG E++

-2446-

Sbjct: 121 ITSRANHWDSILAYQEYGPVIACEDTGDID-VPCAFNDRKTAYAESFRYLKSRGHENIAF 179

Query: 181 LFSRNNESSATYQSALLAYQEVYQGLSSPYMVGNVHDFNDG-LNLSYQLVKEVSIDGIL 239

5 Sbjct: 180 TCVREADRSPSTADKAAAYKAVCGRLDRHMLSG-CNDMNDGELAAEHFYMGRVPTAIY 238

Query: 240 ATSDEVAAGLIKGYEESRKKCPYIIGQECLLVGQLLKLPTIDHKSYYLGKLAFAKQALAEK 299

A SDEVAAG I + + IIG+ + ++L P++D LG AF L ++

10 Sbjct: 239 ANSDEVAAG-IHLFAKKNWVDVEIIGEGNTSISRVLGFPSLDLNLLEQLGIAAFSLFLQDE 297

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2171

15 A DNA sequence (GBSx2288) was identified in *S.agalactiae* <SEQ ID 6707> which encodes the amino acid sequence <SEQ ID 6708>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3451(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAC21682 GB:U32686 conserved hypothetical protein [Haemophilus influenzae Rd]
Identities = 79/264 (29%), Positives = 134/264 (49%), Gaps = 16/264 (6%)

Query: 1 MTIKRIFCDMDGTLNSEQVSKSNATLIREAA---IPVTLVSARAPMEMKDAVDALQLG 57

30 M K +F D +GTLL S+ +S +I+ IP +SAR+P+ + L+

Sbjct: 1 MMVKAVSFDFNGITLLTSQHTISPRTVVVIKRLTANGIFVFPISARSPLGILPYWKQLETN 60

Query: 58 GVQVAFNGGLIYRIGDNNQVLPPIHTQIIKKSTVKQLLRGIRFHFPOVSLSYDNLNNWYCD 117

V VAF+G LI N + PI++ I+ + ++ + H P + ++YY N+ +

35 Sbjct: 61 NVLVAFSGALIL---NQNLPIYSVQIEPKDILEINTVLAEH-PLLGVNYYTNNDCAR 115

Query: 118 KID-EGIRYEHSLTQQCPTFIHNEQFLEGHTNTFKIMMITFDEANMLELEKYLSLELP 176

++ + + YE S+T+ IH D+ T + + I + ++E+E L+ + P

40 Sbjct: 116 DVENKWIYERSVTK---IEIHPFDEVA---TRSPHKIQIIGEAEEIIEIEVLLKE-KFP 168

Query: 177 EITIQRSKGAYLEITHLLAKSKGIAYILQKEQLAREETAAFGDGHNDLPMLEMVGYPIV 236

++I RS +LE+ H A K + ++ + E AFGD NDL MLE VG +

45 Sbjct: 169 HLSICRSHANFLEVMHKSATKGSAAVRFLDYFGVQTNEVIAFGDNFNDLDMLEHVGLGVA 228

Query: 237 MDNAFDDIKAIAYQLTKSNDEEDGV 260

50 M NA ++IK A +T +N+EDG+

Sbjct: 229 MGNAPNEIKQAANVVTATNNEDGL 252

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50 Example 2172

A DNA sequence (GBSx2289) was identified in *S.agalactiae* <SEQ ID 6709> which encodes the amino acid sequence <SEQ ID 6710>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2854(Affirmative) < succ>

-2447-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2173

10 A DNA sequence (GBSx2290) was identified in *S.agalactiae* <SEQ ID 6711> which encodes the amino acid sequence <SEQ ID 6712>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have an uncleavable N-term signal seq

15	INTEGRAL	Likelihood = -10.51	Transmembrane	392 - 408 (376 - 417)
	INTEGRAL	Likelihood = -9.92	Transmembrane	440 - 456 (433 - 461)
	INTEGRAL	Likelihood = -6.42	Transmembrane	52 - 68 (51 - 70)
	INTEGRAL	Likelihood = -6.32	Transmembrane	29 - 45 (9 - 48)
	INTEGRAL	Likelihood = -6.32	Transmembrane	309 - 325 (308 - 328)
	INTEGRAL	Likelihood = -4.46	Transmembrane	12 - 28 (9 - 29)
20	INTEGRAL	Likelihood = -3.29	Transmembrane	463 - 479 (462 - 479)
	INTEGRAL	Likelihood = -2.07	Transmembrane	353 - 369 (352 - 369)
	INTEGRAL	Likelihood = -1.17	Transmembrane	374 - 390 (374 - 390)
	INTEGRAL	Likelihood = -0.85	Transmembrane	247 - 263 (247 - 263)
	INTEGRAL	Likelihood = -0.06	Transmembrane	278 - 294 (278 - 294)

25 ----- Final Results -----

bacterial membrane --- Certainty=0.5203 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23742 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 325/705 (46%), Positives = 478/705 (67%), Gaps = 3/705 (0%)

35	Query: 1	MLQLTKYFPLKPIYLALLVVFQIYLLVFSWTMLGCAFLLFSEFIFLIYQYDRETIFKTIAIV 60
		MLQ K F + IYL+ L+ +Y +FS + L +F + L Q+ ++ K + I
	Sbjct: 1	MLQWIKNFSIPLIYLSFLLWLWLYAIFSASYLALLGFLVCLFIQFPWKSAGKVLIIIC 60
40	Query: 61	IFFLFYFLWQNHNMVQYQRPVNHISQIKVRIDTISINGDVLVSFOADASGNTYQAFYTLK 120
		F F+F++QN + Q + + + + + + DT+ +NGD LSF+ A G +Q +Y L+
	Sbjct: 61	GIFGFWFVFQNWQQSQASQNLADSVRVRILPDTVKVNGDSLSPRGKADGRIFQVYYKLQ 120
45	Query: 121	NKSEKDYFQNLDDNNIMIADIKLEEAEEERRHFNGFDYRQYLKRHGIYRIAKVTIKIQRIL 180
		++ EK+ FQ L + I + KL E E +R+F GF+Y+ YLK GIY+ + KI+ ++
	Sbjct: 121	SEEEKAFQALTDLHEIGLEGKLESEPEGQRNFGGFNYQAYLKTQGIYQTLNIIKIQSLQK 180
50	Query: 181	FQHRFFALMSKWRRAIVISQT-FPNPMRHYMSGLLFGYLDKTFDDMSDLYSSLGIIHL 239
		+S RR A+V +T FP+PMR+YM+GLL G+LD F++M++LYSSLGIIHL
	Sbjct: 181	IGSWDIGENLSSLRKAVVWIKTHFPDPMRNYMTGLLLGHLDTFEEMNELYSSLGIIHL 240
55	Query: 240	FALSGMQVGFFLGIFRYICLRIGRLRDHVWLLQIPFSLIYAGLTGFSISVVRALIQSLLS 299
		FALSGMQVGFF+ F+ + LR+GL + + L PFSLIYAGLTGFS SV+R+L+Q LL+
	Sbjct: 241	FALSGMQVGFFMNGFKKLLRLGLTQEKLKWLTYPFSLIYAGLTGFSASVIRSLQKLLA 300
	Query: 300	HSGVKDENFALCLLICISLPHSLLTGTVLSFAYAFILTMSTFDHFSSIKKVAIESLT 359
		GVK +N AL +L+ I +P+ T GGVLS AYAFILTM S + +K VA ESL
	Sbjct: 301	QHGKGLDNCALTIVLVLFIVMPNFFFTAGGVLS CAYAFILTMPSKEG-EGLKAVASESLV 359
	Query: 360	VSVGILPILTYFSGFQPIISIIITALLSFAFDIIFLELLTVIFVLSPIVKLSCINSLFEI 419
		+S+GILPIL++YF+ FQP SI+LT + SF FD+ FLPLL+++FVLS + + +N +FE

-2448-

Sbjct: 360 ISLGILPILSFYFAEFQPWSILLTFVFSFLFDLTFPLPLLSILFVLSFLYPVQLNFIFEW 419

Query: 420 LEVLLKWTGQLFPRPLIFGKPSLFLLLVMIILGLLYDYHSHKCFRYCSLLIIFTLFFIT 479
LE +++ Q+ RPL+FG+P+ +LLI+++I L L+YD + L+I LF +T

5 Sbjct: 420 LEGIIRLVSQVTSRPLVFGQPNTWLLILLISLALVYDLRKNKKLTVLCCLITGLFLLT 479

Query: 480 KNPITNEVAILDVGQGSILVRDWLGTKILIDTGGRRV- FEQPEEWKQKVNQSNKRTLI 538
K+P+ NE+ +LDVGQG+SI +RD GTKILID GG+ +++ ++W++K+ SNA+R+LI

10 Sbjct: 480 KHPLENEITMLDVGGESIFLDRDVTGKILIDVGGAESYKIKKKWQEKMTTSNAQRS LI 539

Query: 539 PYLKSRGISKIDDLVITHDTDTHMGDMEVISKHFKVARLITSSGSLTNSQYVKHLSKIGV 598
PYLKSRG++KID L++T+TD +H+GD+ ++K F V ++ S SL ++V L

15 Sbjct: 540 PYLKSRGAKIDQLILTNTDKEHVGDLSMTKAFHVGEILVSKDSLKQKEFV AELQATQT 599

Query: 599 AVKSIEAGDKLAVMGSYLQVLYPWHKGDGKNNDISIVLYGHLLGKGFLETFGDLEEEGEKQL 658
V+S+ G+ L + GS L+VL P GDG ++D++VLYG L K FLFTG+LEE+GEK L

20 Sbjct: 600 KVRSMIVGENLPFIGSQLEVLSPRKMGDGGHDDTLVLYGKFLDKQFLFTGNLEEKGEKDL 659

Query: 659 LEAYPNLSVDILKAGHHGSKGSSSLFLLKKLSPSVVLVSAGKNR 703
L+ YP+L V++LKA HG+K SSS +FL+KL P + L+S GK+NR

Sbjct: 660 LKHYPDLKVNVLKASQHGNNKSSSPAFLEKLKPELTLISVGKSNR 704

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6713> which encodes the amino acid sequence <SEQ ID 6714>. Analysis of this protein sequence reveals the following:

25 Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.19	Transmembrane	394 - 410 (380 - 422)
INTEGRAL	Likelihood = -8.28	Transmembrane	54 - 70 (52 - 72)
INTEGRAL	Likelihood = -6.32	Transmembrane	356 - 372 (355 - 377)
30 INTEGRAL	Likelihood = -4.73	Transmembrane	8 - 24 (7 - 25)
INTEGRAL	Likelihood = -4.30	Transmembrane	30 - 46 (29 - 50)
INTEGRAL	Likelihood = -3.88	Transmembrane	249 - 265 (249 - 267)
INTEGRAL	Likelihood = -3.40	Transmembrane	467 - 483 (465 - 484)
INTEGRAL	Likelihood = -2.39	Transmembrane	325 - 341 (325 - 347)
35 INTEGRAL	Likelihood = -0.43	Transmembrane	441 - 457 (441 - 458)

----- Final Results -----

bacterial membrane	---	Certainty=0.5076(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
40 bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GP:AAC23742 GB:AF052208 competence protein [Streptococcus pneumoniae]

Identities = 311/706 (44%), Positives = 458/706 (64%), Gaps = 10/706 (1%)

45 Query: 5 WTKLVPLSKIQFAFLILVFFYQIHSPSWLTFL-LSLSLICLLVKRLSKK--EFLGVFAIL 61
W K + I +FL+L +Y I S S+L L L+CL ++ K + L + I

Sbjct: 4 WIKNFSIPLIYLSFLLLWLYYAIFASYLALG FVFLVCLFIQFPWKSAGKVLIIICGIF 63

50 Query: 62 SFCALFLLYQKQQLVQKLEIQPVQITSVALVPDSIRINGDQLAVLGRHGKHSYQLFYRLK 121
F +F +Q+ Q Q L + V ++PD++++NGD L+ G+ +Q++Y+L+

Sbjct: 64 GFWFVFQNWQQSQASQNLADS---VERVRILPDTVKVNGDSL SFRGKADGRIFQVYKQLQ 120

55 Query: 122 SQAEAQLFKKEHRLVMHAKVTLEKAEVVRNFKGFNYQTFLLTYQGIYRIGKVEQIEQLEV 181
S+ E + F+ + + L + E RNF GFNYQ +L QGIY+ +++I+ L+

Sbjct: 121 SEEEKAEFAQALTDLHEIGLEGKLESEPEGQRNFGGFNYQAYLKTQGIYQTLNKKIQSLQK 180

Query: 182 ISPESICDYLLSSLRRAIVHCQQHFPMPMSHYLTGLLFGYLDKSFSGEMTDYYSQLGIIHL 241
I I + LSSLRRA+V + HFP PM +Y+TGLL G+LD F EM + YS LGIIHL

60 Sbjct: 181 IGSWDIGENLSSLRRAKVVWIKTHFPDPMRNYMTGLLGHLDTFEEMNELYSSLGIIHL 240

Query: 242 FALSGMQVGFFLTFCRRVLLLLAVPLEWIKWIELPFACFYAALTGYSISVIRSLVQSQR 301
FALSGMQVGFF+ F+++LL L + E +KW+ PF+ YA LTG+S SVIRSL+Q L

65 Sbjct: 241 FALSGMQVGFFMNGFKLLLRGLTQEKLKWLTYPFSLIYAGLTGFSASVIRSLQKLLA 300

-2449-

Query: 302 HLGKGLDNLACTFLLVFLWDAHFLMTVGGVLTFSYAFLLTVVTVEELSGAKRQLVQVLT 361
 G+KGLDN A T L++F+ +F T GGVLT+ +YAF+LT+ + +E G K + L
 Sbjct: 301 QHGKGLDNLCALTIVLVLFIIVMPNFEFTAGGVLS CAYAFILTMP S-KEGEGLKAVASESLV 359

5 Query: 362 ISLGILPFLFYFSSFNPMMSVLTGLLSYLFDFILPCLLCLVFCLSPLVTVSICNHLFIL 421
 ISLGILP L FYF+ F P S++LT + S+LFDD LPLL ++F LS L V N +F
 Sbjct: 360 ISLGILPILSFYAFQFWSILLTFVFSFLFDLTFLPILLSILFVLSFLYPVLIQNFIFEW 419

10 Query: 422 LEKVIQFLGNTFNSSLVFGSPTSWHLLILVISFAIFYDYRQ-VRQRVITCGLVIALTLLS 480
 LE +I+ + + LVFG P +W L++L+IS A+ YD R+ +++ + C L+ L LL+
 Sbjct: 420 LEGIIRLVSQVTSRPLVFGQPNWLLILLLISLALVYDLRKNIKLITVLCCLITGLFLLT 479

15 Query: 481 VKYPLTNEVTFIDIGQGSILVREWTGKNLLIDVGGGR-PFFSSKEHWRRGHHVANAQKTL 539
 K+PL NE+T +D+GQG+SI +R+ TGK +LIDVGG+ + + W+ +NAQ++L
 Sbjct: 480 -KHPLENEITMLDVGGQGESIFLRDVTGKTILIDVGGKAESYKKIKKWQEKMTTSNAQRSL 538

20 Query: 540 IPYLKSRGIHTIDQLLVTHADTDHMGDIEVVAKAIRIKEILTSQGSLSHPSFVRRLRLK 599
 IPYLKSRG+ IDQL++T+ D +H+GD+ + KA + EIL S+ SL FV L+ +
 Sbjct: 539 IPYLKSRGVAKIDQLILTNTPKEHVGDLSEMTKAFHVGEILVSKDSLKQKEFVAELQATQ 598

25 Query: 600 CHVRVLAAGDQLEPIMGSVLQVLYPWQLGDGKNNDLSVLVYGRLLNRTFLFTGDLEKEGENE 659
 VR + G+ LPI GS L+VL P ++GDG ++D+LVLYG+ L++ FLFTG+LE++GE +
 Sbjct: 599 TKVRSMIVGENLPFGSQLEVLSPRKMGGGHDDTLVLYGKFLDKQFLFTGNLEEKGEKD 658

30 Query: 660 IIKRYPQLRVLYLKAGHHGNSNTSSSAFLDHIQPKVAFISAGKNR 705
 ++K YP L+V+ LKA HG+ SSS AFL+ ++P++ IS GK+NR
 Sbjct: 659 LLKHPDLKVNVLKASQHGKNKSSSPAFLEKLPKELTLISVGKSNR 704

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 346/743 (46%), Positives = 491/743 (65%), Gaps = 3/743 (0%)

Query: 5 TKYFPLKPIYLALLVFQIYLLVFSWMLGCAFLLESFIFLIYQYDRETIKTIIVIFFL 64
 TK PL I A L+ + + S + L L L+ + ++ AI+ F
 Sbjct: 6 TKLVPLSKIQFAFLILVFFYQIHSPSWLTFLLSLSLICLLVKRLSKKEFLGVFAILSFC 65

35 Query: 65 FYFLWQNHNMNVQYQRPVNHISQIKVRIDTISINGDVLVSFQADASGNTYQAFYTLKNKSE 124
 + L+Q + + + P I+ + + D+I INGD L+ ++YQ FY LK+++E
 Sbjct: 66 LFLLYQKQQLVQKLEIQPVQITSVALVPDSIRINGDQLAVLGRHGKHSYQLFYRLKSQAE 125

40 Query: 125 KDYFQNLNMMIADIKLEAEERRHFNFGDYRQYLRHGIYRIAKVTIKQIRLFQHR 184
 F+ +++ A + LE+AEE R+F GF+Y+ +L GIYRI KV +I+Q+ +
 Sbjct: 126 AQLFKKEHRWLVMAKVTLEKAEVRNFKGFNYQTFLTYQGIYRIGKVEQIEQLEVISPE 185

45 Query: 185 SFFALMSKWRRAIV-ISQTFPNMRHYMSGLLFGYLDKTFDDMSDLYSSLGIIHLFALS 243
 S +S RR AIV Q FP PM HY++GLLFGYLDK+F +M+D YS LGIIHLFALS
 Sbjct: 186 SICDYLSSLRRAIVHCQHFPRPM SHYLTGLLFGYLDKSGFEMTDYYSQLGIIHLFALS 245

50 Query: 244 GMQVGFFLGIFRYICLRIGRLDHVWLLQIPFSLIYAGLTGFSISVVRALIQSLLSHSGV 303
 GMQVGFFL FR + L + + L+ + +++PF+ YA LTG+SISV+R+L+QS L H G+
 Sbjct: 246 GMQVGFFLTCFRRVLLLLAVPLEWIKWIELPFACFYAALTGYSISVIRSLVQSQRHLGI 305

55 Query: 304 KKDENFALCLLICLSLPHSLLTGGVLSFAYAFILTMSTFDHFSSIKKVAIESLTVSVG 363
 K +N A L+ + H L+T GGVLT+F+YAF+LT+ + + S K+ ++ LT+S+G
 Sbjct: 306 KGLDNLACTFLLVFLWDAHFLMTVGGVLTFSYAFLLTVVTVEELSGAKRQLVQVLTISLG 365

60 Query: 364 ILPILTYFSGFQPSISIIILTALLSFAFDIIFLPLLTIVFVLSPIVKLSLSCINSLFEILEVL 423
 ILP L +YFS F P+S++LT LLS+ FD+ LPLL ++F LSP+V +S N LF +LE +
 Sbjct: 366 ILPFLFYFSSFNPMMSVLTGLLSYLFDFILPCLLCLVFCLSPLVTVSICNHLFILLEKV 425

65 Query: 424 LKWTGQLFPRPLIFGKPSLFLLIVMIIILGLLYDYYHSC-FRYCSLLIIFTLFFITKNP 482
 +++ G F L+FG P+ + L++++I + YDY + C L+I TL + K P
 Sbjct: 426 IQFLGNTFNSSLVFGSPTSWHLLILVISFAIFYDYRQVRQRVITCGLVIALTLLSV-KYP 484

Query: 483 ITNEVAILDVGGQGSILVRDWLGTKITLIDTGGRVRFQPEEWKQKVNQSNKRTLIPYLK 542
 +TNEV +D+GQGSILVR+W GK +LID GGR F E W++ + +NA++TLIPYLK
 Sbjct: 485 LTNEVTFIDIGQGSILVREWTGKNLLIDVGGRPFFSSKEHWRRGHHVANAQKTLIPYLK 544

-2450-

Query: 543 SRGISKIDDLVITHDTDTHMGDMDEVISKHFKVARLITSSGSLTNSQYVKHLSKIGVAVKS 602
 SRGI ID L++TH DTDHMGD+EV++K ++ ++TS GSL++ +V+ L ++ V+
 Sbjct: 545 SRGIHTIDQLLVITHADTDHMGDIEVVAKAIRIKEILTSQGSLSHPSFVRRLRLKCHVRV 604

5 Query: 603 IEAGDKLAVMGSYLQVLYPWHKGDGKNNDSIVLYGHLLGKGFLLFTGDLEEEGEKQLLEAY 662
 + AGD+L +MGS LQVLYPW GDGKNNDS+VLYG LL + FLFTGDLE+EGE ++++ Y
 Sbjct: 605 LAAGDQLPIMGSVLQVLYPWQLGDGKNNDSLVLVYGRLLNRTFLFTGDLEKEGENEIKRY 664

10 Query: 663 PNLSDVILKAGHHGSKGSSSLSFLLKLSPSVVLVSAGKNNRYQHPHQETLQRFQKIKSKI 722
 P L VD LKAGHHGS SSS +FL + P V +SAGKNNRYQHPH+ETL R + +
 Sbjct: 665 PQLRVDYLKAGHHGSNTSSSAFLDHIQPKVAFISAGKNNRYQHPHRETLLARLEDRQITY 724

Query: 723 FRTDQSGTIRLTGWKWHIQTVR 745
 +RTD G IRLTG WH++TVR
 15 Sbjct: 725 YRTDTQGAIRLTGRTSWHLETVR 747

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2174

20 A DNA sequence (GBSx2291) was identified in *S.agalactiae* <SEQ ID 6715> which encodes the amino acid sequence <SEQ ID 6716>. This protein is predicted to be competence protein (comEA). Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq
 25 INTEGRAL Likelihood = -3.77 Transmembrane 18 - 34 (14 - 36)

----- Final Results -----
 bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 96/217 (44%), Positives = 138/217 (63%), Gaps = 4/217 (1%)

35 Query: 3 EIVLEKIKSHKWETTGLIIVGLLLFGILGLNHFHGTTHKEDNLNINLEK-KVSTITEKKVP 60
 E ++EKIK +K +GLL+ G L T KE NL + ++EK+V
 Sbjct: 2 EAIIEKIKEYKIIVICTGLGLLVGGFFLLKPAPQTPVKETNLQAEVAAVSKDLVSEKEVN 61

40 Query: 61 MISHVKDKVSNQVTVDVKGAVNHPGVYSLPSQSRVTDIAIKRAGGLSNLADSKSVNLAQKL 120
 + + +TVDVKGAV PG+Y LP SR+ DA+++AGGL+ ADSKS+NLAQK+
 Sbjct: 62 KEEKEEPLEQDLITVDVKGAVKSPGIYDLPVGSRINDAVQKAGGLTEQADSKSLNLAQKV 121

45 Query: 121 QDETVIYVAQKGEKITVVEEEKANNIATQGNKSGKINLNKADLSSIQTTISGVGAKRAQDI 180
 DE ++YV KGE+ V ++ A+ + + K+NLNKA L L+ + G+G KRAQDI
 Sbjct: 122 SDEALVYVPTKGEE--AVSQQTGLGTASSISKEKKVNLNKALEELKQVKGLGKRAQDI 179

Query: 181 LDYRDSQGGFKTIDDLKNVSGIGETLEKLRQDVTID 217
 +D+R++ G FK++D+LK VSGIG KT+EKL+ VT+D
 50 Sbjct: 180 IDHREANGKFKSVDELKKVSGIGGKTIEKLKDYVTVD 216

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6717> which encodes the amino acid sequence <SEQ ID 6718>. Analysis of this protein sequence reveals the following:

Possible site: 36
 55 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.61 Transmembrane 22 - 38 (16 - 42)

----- Final Results -----

-2451-

bacterial membrane --- Certainty=0.4843(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 82/179 (45%), Positives = 124/179 (68%), Gaps = 4/179 (2%)

10 Query: 42 NRQSKAAVPALREISPVKQQVSEEEKIEQEDSSILVDLKGAVQKEGVYKLTASSRVRDVI 101
 N Q++ A + +++ K+ EEK+E E I VD+KGAV+ G+Y L SR+ D +
 Sbjct: 42 NLQAEVAAVS-KDLVSEKEVNKEEKEPLEQDLITVDVKGAVKSPGIYDLVPVGSRINDAV 100

15 Query: 102 ELAGGLTSEADKHAINFAEKLTDQVYVVPKQGEEISVLPRLSVSGKKETASKDQSKVHI 161
 + AGGLT +AD ++N A+K++DE +VYVP +GEE + + G + SK++ KV++
 Sbjct: 101 QKAGGLTEQADSKSLNLAQKVSDEALVYVPTKGEE--AVSQQTGLGTASSISKEK-KVNL 157

20 Query: 162 NKASLEELQHIPPIGAKRAQDIIDMRDKLGGFKALEDLRQVSGIGECTLEKLKDDIFLD 220
 NKASLEEL+ + G+G KRAQDIID R+ G FK+++L++VSGIG KT+EKLD + +D
 Sbjct: 158 NKASLEELKQVKGLGGKRAQDIIDHREANGKFKSVDELKKVSGIGCKTIEKLKDYVTVD 216

An alignment of the GAS and GBS proteins is shown below.

Identities = 81/166 (48%), Positives = 111/166 (66%), Gaps = 10/166 (6%)

25 Query: 62 ISHVKDKVSNQ-----VTVDVKGAVNHPGVYSLPSQSRVTDKIKRAGGLSNLADSK 112
 IS VK +VS + + VD+KGAV GY L + SRV D I+ AGGL++ AD
 Sbjct: 55 ISPVKQQVSEEEKIEQEDSSILVDLKGAVQKEGVYKLTASSRVRDVIELAGGLTSEADKH 114

30 Query: 113 SVNLAQKLQDETVIYVAKGEEKITVVEEEKANNIA-TQNSKGKINLNKADLSSLTISG 171
 ++N A+KL DE V+YV ++GE+I+V+ + T + K+++NKA L LQ I G
 Sbjct: 115 AINFAEKLTDQVYVVPKQGEEISVLPRLSVSGKKETASKDQSKVHINKASLEELQHIPP 174

35 Query: 172 VGAKRAQDILDYRDSQGGFKTIDDLKNVSGIGECTLEKLKQDVTID 217
 +GAKRAQDI+D RD GGFK ++DL+ VSGIGECTLEKL+ D+ +D
 Sbjct: 175 IGAKRAQDIIDMRDKLGGFKALEDLRQVSGIGECTLEKLKDDIFLD 220

A related GBS gene <SEQ ID 8989> and protein <SEQ ID 8990> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 5.70
 GvH: Signal Score (-7.5): -2.58
 Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -3.77 threshold: 0.0
 INTEGRAL Likelihood = -3.77 Transmembrane 18 - 34 (14 - 36)
 45 PERIPHERAL Likelihood = 10.40 73
 modified ALOM score: 1.25

*** Reasoning Step: 3

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

44.3/64.1% over 215aa

Streptococcus

pneumoniae
 GP|3211753| competence protein Insert characterized

60 ORF01930(304 - 951 of 1014)
 GP|3211753|gb|AAC23741.1|AF052208(1 - 216 of 216) competence protein {Streptococcus pneumoniae}

```
%Match = 25.0
%Identity = 44.2 %Similarity = 64.1
Matches = 96 Mismatches = 75 Conservative Sub.s = 43
```

60 Subject: 179 AAMTEALNORIKELNKLDO 198

:2453-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6721> which encodes the amino acid sequence <SEQ ID 6722>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.83    Transmembrane    241 - 257 ( 234 - 266)
    INTEGRAL    Likelihood = -4.41    Transmembrane    27 - 43 ( 26 - 44)

----- Final Results -----
10      bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

15  >GP:CAB12793 GB:Z99109 similar to 1-acylglycerol-3-phosphate
    O-acyltransferase [Bacillus subtilis]
    Identities = 59/198 (29%), Positives = 104/198 (51%), Gaps = 6/198 (3%)

Query: 29  YAYLRGLVVFLLWVVGNAHYHHEEKMLDASENYILVAPHRTFWDVPVMAFAARPKQFIF 88
20  Y +   + +L + G Y+ E L A +++ H + D + + P Q +
Sbjct: 2   YKFCANALKVILSLRGGVKVINKEN--LPADSGFVIACHTSGWVDVITLGVGILPYQIH 59

Query: 89  MAKKELFANRLFPAWIKMCGAFFIDRDKPSPDAIRYPVNMLKKSNNRSLMFPSGSRHSQE 148
25  MAKKELF N+   ++K AFP+DR+ P P +I+ P+ +LK+   + +FPSG+R S++
Sbjct: 60  MAKKELFQNKWIGSFLKKIHAFPDRENPGPSSIKTPIKLLKE-GEIVGIFPSGTRTSED 118

Query: 149 V--KGGVAVIKLAQKIMPAAYQGPMGVKGLLAGERVDMTFGNPIDVSDIKRM-NDEGI 205
      V K G IA++ K ++PAAYQGP S K L ++ + G P+ +D + + E +
30  Sbjct: 119 VPLKRGAVTIAQMGKAPLVPAAYQGPSSGKELFKKKGKMKLIIGEPHQAQFAHLPSKERL 178

Query: 206 AEVANRIQAEFDRIDDEL 223
      A +   +   ++++L
Sbjct: 179 AAMTEALNQRIKELENKL 196

```

35 An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 186/244 (76%), Positives = 212/244 (86%)

Query: 1   MFYTYLRTLVMFLIWVANGNAHYHNEKMLKDDENYILVAPHRTFWDVPVMAFAARPKQF 60
      +FY YLR LV+FL+WV NGNAHYH+E+KML ENYILVAPHRTFWDVPVMAFAARPKQF
40  Sbjct: 27 VFYAYLRGLVVFLLWVVGNAHYHHEEKMLDASENYILVAPHRTFWDVPVMAFAARPKQF 86

Query: 61  IFMAKKELFTNRLFGWIKMCGAFFIDREKPGQDAIRYPVKMLKNSNRSLVMFPSGSRHS 120
      IFMAKKELF NRLF WNIKMGAFPIDR+KP DAIRYPV MLK SNRSL+MFPSGSRHS
45  Sbjct: 87 IFMAKKELFANRLFPAWIKMCGAFFIDRDKPSPDAIRYPVNMLKKSNNRSLMFPSGSRHS 146

Query: 121 KDVKGGVAVIKMAKVRIMPAAYRGPMVFNLLKGHRVDMNFGNPIDVSDIKRMDAEGIA 180
      ++VKGGVAVIAK+AKV+IMPAAY+GPM K LL G RVDM FGNPIDVSDIKRM+ EGIA
50  Sbjct: 147 QEVKGGVAVIKLAQKIMPAAYQGPMGVKGLLAGERVDMTFGNPIDVSDIKRMNDEGIA 206

Query: 181 EVSRRIQEEFDRIDRENETYDDGKKLNPLTYIYRLPLAIIAIVLLVLTILFSYLASFVWD 240
      EV+ RIQ EFDR+D E + GK NPLTY+YRLPL ++ +V+L+LT++FSY+ASFVW+
55  Sbjct: 207 EVANRIQAEFDRIDDELAPFQPGKARNPLTYLYRLPLGLVLVVVLLLTMLFSYIASFVWN 266

Query: 241 PQKH 244
      P KH
Sbjct: 267 PDKH 270

```

60 SEQ ID 6720 (GBS171) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 36 (lane 2; MW 25kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 3; MW 49.8kDa).

-2454-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2176

A DNA sequence (GBSx2293) was identified in *S.agalactiae* <SEQ ID 6723> which encodes the amino acid sequence <SEQ ID 6724>. Analysis of this protein sequence reveals the following:

Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3268(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB11810 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 113/244 (46%), Positives = 173/244 (70%), Gaps = 2/244 (0%)

Query: 6 LKENERIDQLFSTDVKIIQNKEVFSYSIDSVLLSRFPKLP-SRGLIVDLCSENGAVGLFA 64
 L ++ER+D L + D+KIIQ+ VF++S+D+VLLS+F +P +G IVDLC+GNG V L
20 Sbjct: 4 LHDDERLDYLLAEDMKIIQSPTVFAFSLDAVLLSKFAYVPIQKGKIVDLC+TNGGIVPLLL 63

Query: 65 STKTNATIIIEIQLQESLADMAKRSIKLNKLEKQVTMINDDLKNLLDHVQRSNVDLMCNP 124
 ST++ A I+ +E+QE L DMA RS++ NKL+ Q+ +I+DDLKN+ + + + D++ CNP
25 Sbjct: 64 STRSKADILGVEIQERLHDMAVRSVEYNKLDQIQIHHDDLKNMPEKLGHNRYDVVTCNP 123

Query: 125 PYFKASETSKKNLSPHYLLARHEITTNLREICQIAQHALKTKGRIAMVHRPDRFLEIIDT 184
 PYFK + +++N++ H +ARHEI L ++ ++ LK G+ A+VHRP R LEI +
30 Sbjct: 124 PYFKTPKQTEQNMEHLRIARHEIHCTLEDVISVSSKLLKQGGKAALVHRPGRLLLEIFEL 183

Query: 185 MRQFNLA PKRIQFVYPKLGKDNMLLIEAIKDGSTEGMKILPPLVVHQDNGDYTETIFDI 244
 M+ + + PKR+QFVYPK GK+AN +L+E IK G + +KILPPL V+ + +YT+ I I
35 Sbjct: 184 MKAYQIEPKRVQFVYPKQKQKANTILVEGIKGRPD-LKILPPLFVYDEQNEYTKERTI 242

Query: 245 YFGE 248
 +G+
35 Sbjct: 243 LYGD 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6725> which encodes the amino acid sequence <SEQ ID 6726>. Analysis of this protein sequence reveals the following:

40 Possible site: 48
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2183(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 200/257 (77%), Positives = 228/257 (87%), Gaps = 3/257 (1%)

Query: 1 MIDTILKENERIDQLFSTDVKIIQNKEVFSYSIDSVLLSRFPKLP+SRGLIVDLCSENGAV 60
 MI ILKE ERIDQLFS+DV IIQNK+VFSYSIDSVLLSRFPK+PS+GLIVDLCSENGAV
55 Sbjct: 1 MIKAILKEGERIDQLFSSDVGI IQNKDVFSYSIDSVLLSRFPKMPKGLIVDLCSENGAV 60

Query: 61 GLFASTKTNATIIIEIQLQESLADMAKRSIKLNKLEKQVTMINDDLKNLLDHVQRSNVDLM 120
 GLFAST+T A I+E+ELQE LADM +RSI+LN+LE QVTMI DDLKNLL+HV RS VDLN
65 Sbjct: 61 GLFASTRTKAAIVEVELQERLADMGQRSIQLNQLEDQVTMICDDLKNLLNHVPRSGVDLM 120

Query: 121 LCNPPYFKASETSKKNLSPHYLLARHEITTNLREICQIAQHALKTKGRIAMVHRPDRFLE 180

-2455-

```

          LCNPPYFK+ E+SKKN+S HYLLARHE+TTNL EICQ+A+HALK+ GR+AMVHRPDRFLE
Sbjct: 121 LCNPPYFKSHSSKKNVSEHYLLARHEVTTNLEEICQVARHALKSNGRLAMVHRPDRFLE 180

Query: 181 IIDTMRQFNLA PKRIQFVYPKLGKDANMLLIEAIKDGSIEGMKILPPLVVHQDNGDYTET 240
5      IID++R   LAPKR+QFVYPKLGK ANMLLIEAIKDGS EGM ILPPLVVH++NG+YT+
Sbjct: 181 IIDSLRANGLAPKRVQFVYPKLGKSANMLLIEAIKDGSIEGMTILPPLVVHKENG EYTDH 240

Query: 241 IFDIYFGENGK---SHD 254
          IF+IYFG K +HD
10      Sbjct: 241 IFEIYFGAASKGKPNHD 257

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2177

- 15 A DNA sequence (GBSx2294) was identified in *S.galactiae* <SEQ ID 6727> which encodes the amino acid sequence <SEQ ID 6728>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have no N-terminal signal sequence

20      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1512(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25      The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11811 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
Identities = 40/82 (48%), Positives = 63/82 (76%)

Query: 7  YMYVLECS DGLTYGT TTDVKKRLNTHNTGKGAKYTRARLPVKLLYSEAFNSKQ EAMRAE 66
30      + YV++C D + Y GYT D+ +R+ THN GKGAKYT+ R PV+L+++E+F++K+EAM+AE
Sbjct: 7  FFYVVKCKDNSWYAGYTNDLHKRVKTHNDGKGAKYTKVRRPV E LIFABSFSTKREAMQAE 66

Query: 67 ALFKQKTRQAKLTYIKQH KNEQ 88
          FK+ TR+ K YI++ +N +
35      Sbjct: 67 YYPKKLTRKKKELYIEEKRNSK 88

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6729> which encodes the amino acid sequence <SEQ ID 6730>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
40      >>> Seems to have no N-terminal signal sequence

          ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1838(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 60/84 (71%), Positives = 67/84 (79%), Gaps = 1/84 (1%)

50      Query: 6  AYMYVLECS DGLTYGT TTDVKKRLNTHNTGKGAKYTRARLPVKLLYSEAFNSKQ EAMRA 65
          AYMYVLEC D TLYTGYTTD+K+RL THN GKGAKYTR RLPV LLY E F+SK+ AM A
Sbjct: 6  AYMYVLECBVDKTLTYGT TDLKKRLATHNAGKGAKYTRYRLPVSLLYEVFDSKEAAMSA 65

Query: 66 EALF-KQKTRQAKLTYIKQH KNEQ 88
55      EALF K+KTR KL YI H+ E+
Sbjct: 66 EALFKKRKTRSQKLAYIATHQKEK 89

```

-2456-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2178

A DNA sequence (GBSx2295) was identified in *S. agalactiae* <SEQ ID 6731> which encodes the amino acid sequence <SEQ ID 6732>. This protein is predicted to be autoaggregation-mediating protein (deaD). Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2287(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD20136 GB:AF091502 autoaggregation-mediating protein
[Lactobacillus reuteri]
Identities = 289/504 (57%), Positives = 366/504 (72%), Gaps = 18/504 (3%)

Query: 1 MKFTELNLSQDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL 60
MKF+EL LS +L A++++G+ E +PIQE TIP+ LEGKDVIGQAQTGTGKTAAFGLP +
Sbjct: 1 MKFSELGLSDSLKAIKRSGYEEATPIQEQTIPMVLEGKDVIGQAQTGTGKTAAFGLPII 60

Query: 61 NKIHTEDNTIQAIIAPITRELAVQSQEELFRFRGDKGVKVRVYGGSSIEKQIKALRSGA 120
+ TE+ IQA+II+PTRELA+Q+QEEL+R G+DK V+V+ VYGG+ I +QIK+L+
Sbjct: 61 ENVDTENPNIAIIISPTRELAIQQEELYRLGDKKHVRVQVYGGADIRRQIKSLKQHP 120

Query: 121 HVVVGTPGRLLDLIKRKALKLNHIETLILDEADEMLNMGFLEDIEAIIISRPETRQTLLF 180
++VGTTPGR L D I R +KL+HI+TL+LDEADEMLNMGFLEDIE+II P+ RQTLLF
Sbjct: 121 QILVGTPGRLRDHINRHTVKLDHIKTLVLDEADEMLNMGFLEDIEBIIKETPDDRQTLLF 180

Query: 181 SATMPDPIKRIGVKFMKDPFHVKIKATELTNNVNDQYYVVRVKENEKFDTMTRLMDVDQPE 240
SATMP IKRIGV+FM DPE V+IKA ELT VDQYYVR ++ EKFD MTRL+DV P+
Sbjct: 181 SATMPPEIKRIGVQFMSDPETVRIKAKELTTDLVDQYYVRARDYEFKDIMTRLIDVQDPD 240

Query: 241 LSIVFGRTKRRVDELTRGLKLRGFRAEGIHGDLQNKRLRVIRDFKNDHIDILVATDVAA 300
L+IVFGRTKRRVDEL++GL RG+ A GIHGDL Q+KR +++ FKN+ +DILVATDVAA
Sbjct: 241 LTIVFGRTKRRVDELSKGLIARGYNAAGIHGDLTQDKRSKIMWKFNNELDILVATDVAA 300

Query: 301 RGLDISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSGQSITFVSPNEMGYLTIIENLTKK 360
RGLDISGVTHVYNYDIP DP+SYVHRIGRTGRAG G S+TFV+PNEM YL IE LT+
Sbjct: 301 RGLDISGVTHVYNYDIPSDPSYVHRIGRTGRAGHHGVSLTFVTPNEMDYLHEIEKLTRV 360

Query: 361 RMTGMPATASEAFQAKKKVALKRIARDFED-QELVSK--FDKFKADALELATQYTPEEL 417
RM +KP TA EAF+ ++A F D EL+++ D+++ A +L + +L
Sbjct: 361 RMLPLKPPTABEAFKG-----QVASAFNDIDELIAQDSTDYEEAAEKLLETHNATDL 413

Query: 418 ALYVLSLTVQDPESLPEVEITREKPLPFKPSGGGFKGKGGRGNRGD--RRRNDRGDRR 475
+L+ ++ S V+IT E+PLP + G R N GG+ RR+N R +
Sbjct: 414 VAALLNNMTKEAASEVFKITPERPLPRRNKRN--NRNGNRNNSHGGNHYRRKNFRRHQ 471

Query: 476 GNRDRDRG----SRCDFKRRDDK 495
G+ D+ G SR F R K
Sbjct: 472 GSHRNDNHGKSHSSSRHSFNIRHRK 495

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6733> which encodes the amino acid sequence <SEQ ID 6734>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

-2457-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1108(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 430/545 (78%), Positives = 463/545 (84%), Gaps = 24/545 (4%)

10

Query: 1 MKFTELNLSQDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL 60
 +KFTE NLSQDI SAV AGF + SPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL
 Sbjct: 1 LKFTEFNLSQDIQSAVVTAGFEKASPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL 60

15

Query: 61 NKIHTEDNTIQALIIAPTRELAVQSQEELFRFRGRDKGVKVRVSVYGGSSIEKQIKALRSGA 120
 NKI T +N IQAL+IAPTRELAVQSQEELFRFRGR+KGVKVRVSVYGGSSIEKQIKAL+SGA
 Sbjct: 61 NKI RTNENIIQALVIAPTRELAVQSQEELFRFRGRKGVKVRVSVYGGSSIEKQIKALKSGA 120

20

Query: 121 HVVVGTPGRLLDLIKRKALKLNHIETLILDEADEMLNMGFLEDIEAII SRVPETRQTLLF 180
 H+VVGTPGRLLDLIKRKAL L+H+ETLILDEADEMLNMGFLEDIEAII SRVP RQTLLF
 Sbjct: 121 HIVVGTPGRLLDLIKRKALILDHVETLILDEADEMLNMGFLEDIEAII SRVPADRQTLLF 180

25

Query: 181 SATMPDPIKRGVFKMKDPEHV KIKATELTNNVNDQYYVRVKENEFDTMTRLMVDVQPE 240
 SATMP PIK+IGVFKMKDPEHV+IK ELTNNVNDQYYVRVKE EKFDTMTRLMDV+QPE
 Sbjct: 181 SATMPAPIKQIGVFKMKDPEHVQIKNKELTNNVNDQYYVRVKEQEKFDTMTRLMDVNQPE 240

30

Query: 241 LSIVFGRTRRRVDELTRGLKLRGFRAEGIHGDLQNKRLRVIRDFKNDHIDILVATDVAA 300
 LSIVFGRTRRRVDE+TRGLKLRGFRAEGIHGDLQNKRLRVIRDFKND IDILVATDVAA
 Sbjct: 241 LSIVFGRTRRRVDEITRGLKLRGFRAEGIHGDLQNKRLRVIRDFKNDQIDILVATDVAA 300

35

Query: 301 RGLDISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSGSITFVSPNEMGYLTIENLTKK 360
 RGLDISGVTHVYNYDI QDPESYVHRIGRTGRAGKSG+SITFVSPNEMGYL++IENLTKK
 Sbjct: 301 RGLDISGVTHVYNYDITQDPESYVHRIGRTGRAGKSGSITFVSPNEMGYLSMIENLTKK 360

40

Query: 361 RMTGMKPATASEAFQAKKKVALKRIARDFEDQELVSKFDKFKADALELATQYTPPEELALY 420
 +M ++PATA EAFQAKKKVALK+I RDF D+ + S FDKFK DA++LA ++TPEELALY
 Sbjct: 361 QMKPLRPATAEEAFQAKKKVALKKIERDFADETIRSNFDKFKGDAVQLAAEFTPEELALY 420

45

Query: 421 VLSLTVQDPESLPEVEITREKPLPFKPSGGG---KGKGGRG--NGRGGDRRRNRDRGDR- 474
 +LSLTVQDP+SLPEVEI REKPLPFK GGG GKGGRG N GDRR RGDR
 Sbjct: 421 ILSLTVQDPDSLPEVEIAREKPLPFKYVGGGHGKNGKGGRGDRNRNRDRRGGRGDRN 480

50

Query: 475 -----RGNRDRDDRGSRCDFKRRDDKFKKDNRRQENKKPHKNTSSEKQTGFVI 522
 R RD D DFKR+ + KD +E K SS K TGFVI
 Sbjct: 481 RDERDGRRRQKRDKRDGHGSGNRDFKRSKRNSKDFNKEKK-----SSAKNTGFVI 534

55

Query: 523 RNKGD 527
 R+KG+
 Sbjct: 535 RHKGE 539

A related GBS gene <SEQ ID 8991> and protein <SEQ ID 8992> were also identified. Analysis of this protein sequence reveals the following:

RGD motif 471-473

The protein has homology with the following sequences in the databases:

58.9/74.7% over 494aa

60

GP|4409804| autoaggregation-mediating protein Insert characterized *Lactobacillus reuteri*
 ORF01926(301 - 1785 of 2184)
 GP|4409804|gb|AAD20136.1|AF091502(1 - 495 of 497) autoaggregation-mediating protein
 {*Lactobacillus reuteri*}
 %Match = 37.3
 %Identity = 58.8 %Similarity = 74.6
 Matches = 290 Mismatches = 118 Conservative Sub.s = 78

42 72 102 132 162 192 222 252

-2458-

```

IRHYITKEIPSEAAVAF*IDKL*TLLLYRWVFIAFFLFSEATNRTSNL*KRVIY*IDLILYLFNFNCVTL SRLSEKITN

282      312      342      372      402      432      462      492
KGS*GSFALSFRKEKHLKFTLNLSQDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTLNKIH
5      :||:| | | :| |:::| :| :||| | | | :| | | | | | | | | | | | | | | : :
      MKFSELGLSDSLLKAIKRSGYEATPIQEQTIPMVLEGKDVIGQAQTGTGKTAAFGLPIIENV
      10      20      30      40      50      60

522      552      582      612      642      672      702      732
TEDNTIQALIIAPTRELAVQSQEELFRFRGDKGVKRSVYGGSSIEKQIKALRSGAHVVVGTPGRLLDLIKRKALKLNHI
10      ||: |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      TENPNIIQAIISPTRELAIQTQBELYRLGKDKHVRVQVYGGADIRRQIKSLKQHPQILVGTTPGRLRDHINRHTVKLDHI
      80      90      100      110      120      130      140

15      762      792      822      852      882      912      942      972
ETLILDEADEMLNMGFLEDIEAII SRVPETRTLLFSATMPDPIKRIGVKFMKDPEHVKIKATELTNVNVDQYYVRVKEN
      :||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      KTVLDEADEMLNMGFLEDIESIIKETPDDRQTLTLLFSATMPPEIKRIGVQFMSDPETVRIKAKELTTDLVDQYYVRARDY
      160      170      180      190      200      210      220

20      1002      1032      1062      1092      1122      1152      1182      1212
EKFPDITMRLMDVDQPELSIVFGRTRKRVDELTRGLKLRGFRAEGIHGDLQNKRLRVIRDFKNDHIDILVATDVAARGLD
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      EKFPDITRLIDVQDPDLTI VFGRTKRRVDELSKGLIARGYNAAGIHGDLTQDKRSKIMWKFKNNELDILVATDVAARGLD
      240      250      260      270      280      290      300

25      1242      1272      1302      1332      1362      1392      1422      1452
ISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSGQSITFVSPNEMGYLTIIENLTKKRMTGMKPATASEAFQAKKKVALKR
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      ISGVTHVYNYDIPSDPDSYVHRIGRTGRAGHHGVSITFTVFNEMDYLHEIEKLTRVRMLPLKPPTAEAEAF--KGQVA--
      320      330      340      350      360      370

30      1479      1503      1533      1563      1593      1623      1653      1683
IARDFED-QELVSK--FDKFKADALELATQYTPPEELALYVLSLTVQDPESLPEVEITREKPLPFKPSGGGFGKGGGRNG
      | | ||::| |::| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
35      --SAFNIDIDELAQDSTDRYEEAAEKLLLETHNATDLVAALLNNMTKEAASEVPVKITPERPLPRNRKNRNRNGN--RNNS
      390      400      410      420      430      440      450

1707      1737      1755      1785      1815      1845      1875      1905
RGGD--RRNRDRGDRGRNRDRDRG---SRCDFKRDDKFKKDNRRQENKKPHKNTSSEKQTGFVIRNKGDK*EDYEKG
40      ||: ||:| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
      HGGNHYRRKNFRRHQHGSHRDNHGKSHSSRHSFNIRHRKEN
      470      480      490

```

45 There is also homology to SEQ ID 4454.

SEQ ID 8992 (GBS307) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 7; MW 62kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 2; MW 86.7kDa).

The GBS307-GST fusion product was purified (Figure 208, lane 9; Figure 225, lane 10-11) and used to immunise mice. The resulting antiserum was used for FACS (Figure 272), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2179

55 A DNA sequence (GBSx2296) was identified in *S.agalactiae* <SEQ ID 6735> which encodes the amino acid sequence <SEQ ID 6736>. This protein is predicted to be outer membrane protein (yaeC). Analysis of this protein sequence reveals the following:

-2459-

Possible site: 19

>>> May be a lipoprotein

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB73036 GB:AL139076 putative periplasmic protein [Campylobacter
 jejuni]
 Identities = 89/237 (37%), Positives = 132/237 (55%), Gaps = 3/237 (1%)

 Query: 40 ITVATYSKPTSTFLDLVKDNVKEKGYTLKVVMVSDYIQANIALENKEHDANLLQHEFFMS 99
 IT+ P + L+L+KD+ K KGY LK+V SDYI N ALE KE DANL QH+ F+
 15 Sbjct: 23 ITIGATPNPFGSILLELMKDDFKNGYELKIVEFSDYILPNRALEEKELDANLYQHKKPFLE 82

 Query: 100 IFNKENDGHLVSIPTIYHSLAGFYGQHLKNIAELKDGAKVAIPSDPANMTRALLLQEKK 159
 +N + +L++ TP+ + G Y + +KN+ LK+GA+VAIP+D N +RAL LL++ K
 20 Sbjct: 83 EYNLKKGSNLIATTPVLIAPVGVSYSKKIKNLENLKEGARVAIPNDATNESRALELLEKAK 142

 Query: 160 LITLKNSTSKTKAIEDIITNPKKLRIEPAVALLNLNQAYFEVDLVFNFGYVTKINLVPKR 219
 LI L + KT DI NPKKL+ + L +A + D+ + L P +
 25 Sbjct: 143 LIELNKNITLKTPL--DINKNPKKLKFIELKAAQLPRALDDVDIAIINSNFALGAGLNPSK 200

 Query: 220 DRLLYEKKPDIRFAGALVAREDNKNSDKIKVLKEVLTSKEIRHYITKEIPSEAAVAF 276
 D + E K + + +V R + KNS+K KV+ E+L S + + I + AF
 30 Sbjct: 201 DTIFREDK-NSPYVNYVVVRSEGNSEKTKVIDEILRSDFKFAINEHYKDILIPAF 256

30 SEQ ID 6736 (GBS126) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 7; MW 32kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2180

35 A DNA sequence (GBSx2297) was identified in *S.agalactiae* <SEQ ID 6737> which encodes the amino acid sequence <SEQ ID 6738>. This protein is predicted to be probable permease of ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood = -11.99 Transmembrane 190 - 206 (187 - 215)
 INTEGRAL Likelihood = -8.44 Transmembrane 25 - 41 (16 - 45)
 INTEGRAL Likelihood = -6.48 Transmembrane 69 - 85 (68 - 90)
 INTEGRAL Likelihood = -3.77 Transmembrane 90 - 106 (88 - 109)
 45 INTEGRAL Likelihood = -1.44 Transmembrane 145 - 161 (145 - 161)

----- Final Results -----

50 bacterial membrane --- Certainty=0.5798 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAG08889 GB:AE004963 probable permease of ABC transporter
 [Pseudomonas aeruginosa]
 Identities = 80/206 (38%), Positives = 127/206 (60%), Gaps = 4/206 (1%)

 Query: 15 SFWEITNLMLGLTLILCFLIAFPTGILLFSLRKSYSYLIKHSLAYQLLNFLGLTLRSVPFLIF 74
 +FW MLG +L+ ++ P G+LLF + + Y LL+L + LRS+PF+I
 Sbjct: 24 TFW----MLGGSLLFTVVLGLPLGVLLFLTGPQMFQKAVYTTLSLVVNILRSLPFIIL 79

-2460-

Query: 75 IFILIPLNRLIFGTSFGTIAAILPLTLVSVSLYARYVEQALLNIPQVVDRALSLGANKR 134
 + ++IPL LI GTS G AI PL + + +AR VE AL + + +++ ++GA+ R
 Sbjct: 80 LIVMIPLTLITGTSLGVAGAIPLVVGATPFFARLVETALREV DKGII EATQAMGASTR 139

Query: 135 QIIYYFLIPSIKIDLVLSFTATAISILGYSTIMGVIGAGGLGEYAYRFGYQEYDYPVMYL 194
 QII+ L+P + ++ + T TAI+++ Y+ + GV+GAGGLG+ A RFGYQ + VM +
 Sbjct: 140 QIIWNALLPEARPGIIAAITVTAITLVSYTAMAGVVGAGGLGDLAIRFGYQRFQTDVMVV 199

Query: 195 IVVLFIIYVFILQSLGYFIANRYSRK 220
 VV+ +I V ILQ++G + +SRK
 Sbjct: 200 TVVMLLILVQILQTVGDKLVVHFSRK 225

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2181

A DNA sequence (GBSx2298) was identified in *S.agalactiae* <SEQ ID 6739> which encodes the amino acid sequence <SEQ ID 6740>. This protein is predicted to be ABC transporter, ATP-binding protein (oppF). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5454(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9333> which encodes amino acid sequence <SEQ ID 9334> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22280 GB:U32744 ABC transporter, ATP-binding protein
 [Haemophilus influenzae Rd]
 Identities = 62/174 (35%), Positives = 104/174 (59%), Gaps = 2/174 (1%)

Query: 1 MKMINGLIPYDKGNIYYQGKEVKSFSNDKLRQMRKDIAIYIFQNHNLLAGESVYYHLALVY 60
 ++ +N L G++ G E+ SD +L R+ I IFQ+ NLL+ +V+ ++AL
 Sbjct: 48 IRCVNLLEKPTSGSVIVDGVLTKLSDRELVIARRQIGMIFQHFNLSSRTVFENVALPL 107

Query: 61 KLNHQKVN--HDAINDILDFLGLMDLKQVKCHSLSGGQQQKVAIAMAVLQPKLILCDEI 118
 +L + + I +LD +GL + + +LSGGQ+Q+VAIA A+ PK++LCDE
 Sbjct: 108 ELESESKAKIQEKITALLDLVGLSEKRDAYSNNLSGGQKQRVAIARALASDPKVLLCDEA 167

Query: 119 SSALDTNSEKEIFNLLSDLRKYGISILMIAHLSLLKQYCDRVMILDHQTIVD 172
 +SALD + + I LL ++ GI+IL+I H + ++KQ CD+V ++D +V+
 Sbjct: 168 TSALDEATTQSILKLLKEINRTLGITILLITHEMEVVKQICDQVAVIDQGRIVE 221

There is also homology to SEQ ID 76.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2182

A DNA sequence (GBSx2299) was identified in *S.agalactiae* <SEQ ID 6741> which encodes the amino acid sequence <SEQ ID 6742>. Analysis of this protein sequence reveals the following:

Possible site: 21

-2461-

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2183

A DNA sequence (GBSx2300) was identified in *S.agalactiae* <SEQ ID 6743> which encodes the amino acid sequence <SEQ ID 6744>. Analysis of this protein sequence reveals the following:

15 Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.0904(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9741> which encodes amino acid sequence <SEQ ID 9742> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB87515 GB:AF034138 unknown [Bacillus subtilis]
 Identities = 74/125 (59%), Positives = 92/125 (73%)

30 Query: 5 MGIFSGLMGNASQMDTDKVENQLSDILISDEQVDLAYTLIRDILIVFTNYRLILVDKQGVT 64
 MG GL+GNAS + T V+ +L+ IL+ E+V+ A+ L+RDLIVFT+ RLILVDKQG+T
 Sbjct: 1 MGFIDGLLGNASTLSTA AVQEELAHILLEGEKVEAAFKLVRLIVFTDKRLILVDKQGKIT 60

35 Query: 65 GKKVSYNSIPYASISRFTVETSGHFDLDAELKIWISSAIEPAEVLQFKNDRNIVSIQKAL 124
 GKK + SIPY SISRF+VET+G FDL+ELKIWI A PA QFK D +I IQK L
 Sbjct: 61 GKKTEFQSIPYKSISRFSVETAGRFDL+SELKIWISGAELPAVSKQFKKDESIYDIQKVL 120

Query: 125 ATAVL 129
 A +

40 Sbjct: 121 AAVCM 125

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2184

45 A DNA sequence (GBSx2301) was identified in *S.agalactiae* <SEQ ID 6745> which encodes the amino acid sequence <SEQ ID 6746>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.0921(Affirmative) < succ>

-2462-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9331> which encodes amino acid sequence <SEQ ID 9332> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA74739 GB:Y14370 peptide chain release factor 3
 [Staphylococcus aureus]
 Identities = 274/462 (59%), Positives = 349/462 (75%), Gaps = 9/462 (1%)

Query: 1 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAMVVDSAKGI 60
 M +E++RGISVTSSVMQFDY +NILDTPGHEDFSEDYRTLMAVD+AVMV+D AKG+
 Sbjct: 57 MKVEQERGISVTSSVMQFDYDDYEINILDTPGHEDFSEDYRTLMAVDSAVMVIDCAKGV 116

Query: 61 EAQTKKLFVVKHRNIPVFTTFINKLDRDGREPLDLLEEEVLGIASYPMNWPIGMGKSF 120
 E T KLF+V K R IP+FTTFINKLDR G+EP +LL+E+EE L I +YPMNWPIMG+SF
 Sbjct: 117 EPTTKLKFVKCKMRGIPIFTTFINKLDRVGKEPFELLDEIETLNIEITYPMNWPIMGQSF 176

Query: 121 EGLYDLHNKRLKELYKDERFASIEDG-----DQLFANNPFYEQVKEDIELLQEAGNDFSE 175
 G+ D +K +E ++ +E + D D N+ +EQ E++ L++EAG F
 Sbjct: 177 FGIIDRKSKTIEPRDEENILHLNDDFELEDHAITNDSDFEQAIIEELMLVEEAGEAFDN 236

Query: 176 QAILDGLTPVFFFGSALTNFGVQTFDLTFLEFAPEPHGHKTTEGNVIDPLAKDFSGFVFK 235
 A+L GDLTPVFFFGSAL NFGVQ FL+ +++FAP P+ +T E + P FSGF+FK
 Sbjct: 237 DALLSGDLTPVFFFGSALANFGVQNFLNAYVDFAPMPNARQTKENVEVSPFDDSFSGFIFK 296

Query: 236 IQANMDPRHRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAES-RENVNNAVA 294
 IQANMDP+HRDRIAF+R+VSG FER + + L +K S+V + + ++ ++ V +AVA
 Sbjct: 297 IQANMDPKHRDRIAFMRVVSGAFER-VWMLLCNVLIKSKRSHVQRHLWQTIKKLVNHAVA 355

Query: 295 GDIIGVYDTGTQVQGDITLVGKNKFEEFELPTFTPELFMKVSAKNVMKQKSFHKGIEQLV 354
 GDIIIG+YDTG YQ+GDITL GK + F+ LP FTPE+FMKVSAKNVMKQK FHKGIEQLV
 Sbjct: 356 GDIIGLYDTGNYQIGDITLVGGKQTYSFQDLFPQFTPEIFMKVSAKNVMKQKHFKGIEQLV 415

Query: 355 QEGAIQLYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTMGKKTVRW--INSDD 412
 QEGAIQ YK T + +LGAVGQLQFEVF+HRM+ EYN +VVM P+G+K RW N D
 Sbjct: 416 QEGAIQYYKTLHTNQIILGAVGQLQFEVFEHRMKNEYNVDDVVMPEVGRKIARWDIENEDQ 475

Query: 413 LDERMSSSRNILAKDRFDQPVFLFENDFALRWFADKYPDVKL 454
 + ++M++SR+IL KDR+D VFLFEN+FA RWF +K+P++KL
 Sbjct: 476 ITDKMNTSRNILVKDRYDDLFLFENEFAIRWFEEKFPEIKL 517

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6747> which encodes the amino acid sequence <SEQ ID 6748>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2070(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 447/458 (97%), Positives = 455/458 (98%)

Query: 1 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAMVVDSAKGI 60
 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAMVVDSAKGI
 Sbjct: 57 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAMVVDSAKGI 116

Query: 61 EAQTKKLFVVKHRNIPVFTTFINKLDRDGREPLDLLEEEVLGIASYPMNWPIGMGKSF 120
 EAQTKKLFVVKHRNIPVFTTFINKLDRDGREPL+LLEEEVLGIASYPMNWPIGMG++F
 Sbjct: 117 EAQTKKLFVVKHRNIPVFTTFINKLDRDGREPLELLEEEVLGIASYPMNWPIGMGRAF 176

-2463-

Query: 121 EGLYDLHNKRLLEYKGDERSASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD 180
 EGLYDLHNKRLLEYKGDERSASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD
 Sbjct: 177 EGLYDLHNKRLLEYKGDERSASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD 236

Query: 181 GDLTPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNVIDPLAKDFSGFVFKIQANM 240
 GDLTPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNV+DPLAKDFSGFVFKIQANM
 Sbjct: 237 GDLTPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNVVDPLAKDFSGFVFKIQANM 296

Query: 241 DPRHRDRIAFVRIVSGEGERGMGVNLTRTGKGAKLSNVTQFMAESRENVNAVAGDIIGV 300
 DP+HRDRIAFVRIVSGEGERGMGVNLTRTGKGAKLSNVTQFMAESRENVNAVAGDIIGV
 Sbjct: 297 DPKHRDRIAFVRIVSGEGERGMGVNLTRTGKGAKLSNVTQFMAESRENVNAVAGDIIGV 356

Query: 301 YDTGTQVQGDITLVGKNKFEFEPLPTFTPELFMKVSAKNVMKQKSFHKGIEQLVQEGAIQ 360
 YDTGTQVQGDITLVGKNKFEFEPLPTFTPE+FMKVS KNVMKQKSFHKGIEQLVQEGAIQ
 Sbjct: 357 YDTGTQVQGDITLVGKNKFEFEPLPTFTPEIFMKVSPKNVMKQKSFHKGIEQLVQEGAIQ 416

Query: 361 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTMGKKTVRWINSDDLDERMSSS 420
 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTMGKKTVRWI+ DDL+RMSSS
 Sbjct: 417 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTMGKKTVRWISEDDLQRMSSS 476

Query: 421 RNILAKDRFDQPVFLFENDFALRWFADKYPDVKLEEK 458
 RNILAKDRFDQPVFLFENDFALRWFADKYPDV LEEKM
 Sbjct: 477 RNILAKDRFDQPVFLFENDFALRWFADKYPDVTLEEK 514

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2185

A DNA sequence (GBSx2302) was identified in *Sagalactiae* <SEQ ID 6749> which encodes the amino acid sequence <SEQ ID 6750>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3061(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC38046 GB:AF000954 No definition line found [Streptococcus mutans]
 Identities = 122/142 (85%), Positives = 138/142 (96%)

Query: 1 MLEFAAQKTGKENKEMAVTFVTNERSHELNLKYRDTNRPTDVISLEYKPEVDISFDEEDL 60
 +LEFAAQKTGKE+KEMAVTFVTNERSHELNL+YRDT+RPTDVISLEYKPE +SFDEEDL
 Sbjct: 23 ILEFAAQKTGKEDKEMAVTFVTNERSHELNLKYRDTNRPTDVISLEYKPESLSFDEEDL 82

Query: 61 AENPELAEMLEDFDSYIGELFISIDKAKEQAEYGHSEYEREMGFLAVHGFLHINGYDHYT 120
 A++P+LAE+L +FD+YIGELFIS+DKA+EQA+EYGHSEYEREMGFLAVHGFLHINGYDHYT
 Sbjct: 83 ADDPDLAELVTEFDAYIGELFISVDKAREQAEYGHSEYEREMGFLAVHGFLHINGYDHYT 142

Query: 121 PEEKEMFSLQEEILTAYGLKR 142
 P+EEKEMFSLQEEIL AYGLKR
 Sbjct: 143 PQEEKEMFSLQEEILDAYGLKR 164

There is also homology to SEQ ID 120.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2464-

Example 2186

A DNA sequence (GBSx2303) was identified in *S.agalactiae* <SEQ ID 6751> which encodes the amino acid sequence <SEQ ID 6752>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -15.39    Transmembrane    108 - 124 ( 100 - 131)
    INTEGRAL    Likelihood = -8.92     Transmembrane    61 - 77 ( 52 - 82)
    INTEGRAL    Likelihood = -5.36     Transmembrane    41 - 57 ( 40 - 60)

10 ----- Final Results -----
        bacterial membrane --- Certainty=0.7156(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.
    >GP:AAC38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]
    Identities = 107/133 (80%), Positives = 121/133 (90%), Gaps = 2/133 (1%)

    Query: 1 MDLNDN--NHKKWKNR'LTSSMEFAVTGIFTAFKEERNMRKHLVSAILVILAGLTFQVSM 58
20 MDL DN + KKWKNR'LTSS+EFA+TGIFTAFKEERNM+KH VSA+L ++AGL F+VS+
    Sbjct: 3 MDLRDNKQSQKKWKNR'LTSSLEFALTGIFTAFKEERNMKKHAVSALLAVIAGLVFKVSV 62

    Query: 59 VEWLFLLLSIFLVITFEIINSIAIENVVDLASNYHFSMLAKNAKDMAAGAVLVVSLFAVLV 118
    +EWLFLLLSIFLVITFEI+NSAIENVVDLAS+YHFSMLAKNAKDMAAGAVLV+S FA L
25 Sbjct: 63 IEWLFLLLSIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLVISGFAALT 122

    Query: 119 GLIIFIPKILALL 131
    GLIIF+PKI LL
30 Sbjct: 123 GLIIFVPKIWFL 135

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6753> which encodes the amino acid sequence <SEQ ID 6754>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
35 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -10.67    Transmembrane    63 - 79 ( 41 - 84)
    INTEGRAL    Likelihood = -7.32     Transmembrane    110 - 126 ( 105 - 129)
    INTEGRAL    Likelihood = -5.41     Transmembrane    43 - 59 ( 41 - 62)

40 ----- Final Results -----
        bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

45 >GP:AAC38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]
    Identities = 104/135 (77%), Positives = 119/135 (88%)

    Query: 1 MALHDNNTTKRKWKNR'LTSSLEFALTGVFTAFKEERNLRSHLLSACLACVAGLFFSISA 60
50 M L DN +++KWKNR'LTSSLEFALTG+FTAFKEERN++ H +SA LA +AGL F +S
    Sbjct: 3 MDLRDNKQSQKKWKNR'LTSSLEFALTGIFTAFKEERNMKKHAVSALLAVIAGLVFKVSV 62

    Query: 61 IEWLFLLLAIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLMISGYAVLT 120
    IEWLFLLL+IFLVIT EIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVL+ISG+A LT
55 Sbjct: 63 IEWLFLLLSIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLVISGFAALT 122

    Query: 121 GLIIFIPKIWNIFVH 135
    GLIIF+PKIW + H
60 Sbjct: 123 GLIIFVPKIWFLLFH 137

```

An alignment of the GAS and GBS proteins is shown below.

-2465-

Identities = 98/129 (75%), Positives = 115/129 (88%), Gaps = 2/129 (1%)

Query: 1 MDLNDNN--HKWKNTLTSSMEFAVTGIFTAFKEERNMRKHLVSAILVILAGLTFQVSM 58
 M L+DNN +KWKNT+TSS+EFA+TG+FTAFKEERN+R HL+SA L +AGL F +S
 5 Sbjct: 1 MALHDNNTTKRKWKNTTITSSLEFALTGVFTAFKEERNLRSHLLSACLACVAGLFFSISA 60

Query: 59 VEWLFLLLSIFLVITTFEIIINSATENVVDLASNYHFSMLAKNAKDMAAGAVLVVSLFAVLV 118
 +EWLFLLL+IFLVIT EI+NSATENVVDLAS+YHFSMLAKNAKDMAAGAVL++S +AVL
 10 Sbjct: 61 IEWLFLLLAIFLVITTEIVNSATENVVDLASDYHFSMLAKNAKDMAAGAVLMISGYAVLT 120

Query: 119 GLIIFIPKI 127
 GLIIFIPKI
 Sbjct: 121 GLIIFIPKI 129

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2187

A DNA sequence (GBSx2304) was identified in *S.agalactiae* <SEQ ID 6755> which encodes the amino acid sequence <SEQ ID 6756>. This protein is predicted to be GTPase Era (era). Analysis of this protein
 20 sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.1871(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10017> which encodes amino acid sequence <SEQ ID
 30 10018> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD41632 GB:AF072811 GTPase Era [Streptococcus pneumoniae]
 Identities = 273/299 (91%), Positives = 290/299 (96%)

35 Query: 16 MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQITRNKIMGIYTTTETEQIVFIDT 75
 MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQITRNKIMGIYTT+ EQIVFIDT
 Sbjct: 1 MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQITRNKIMGIYTTDKBQIVFIDT 60

40 Query: 76 PGIHKPKTALGDFMVESAYSTLREVETVLFMVPADKRGKGDMMIIRLKAAPVILVI 135
 PGIHKPKTALGDFMVESAYSTLREV+TVLFMVPADKRGKGDMMIIRLKAAPVILV+
 Sbjct: 61 PGIHKPKTALGDFMVESAYSTLREVDTVLFMVPADKRGKGDMMIIRLKAAPVILVV 120

45 Query: 136 NKIDKVHPDQLLEQIDDFRQMDFKFVVPISALQGNVPTLIKLLTDNLEEGFQYFPDQ 195
 NKIDKVHPDQLL QIDDFR+QMDFKF+VPISALQGNV L+ +L++NL+EGFQYFP DQ
 Sbjct: 121 NKIDKVHPDQLLSQIDDFRNQMDFKFVVPISALQGNVSRVLDLSENLEEGFQYFPDQ 180

50 Query: 196 ITDHPERFLVSEMVREKVLHLLTQQEVPHSVAVVVSMSKRDEETDKVHIRATIMVERDSQK 255
 ITDHPERFLVSEMVREKVLHLLT++E+PHSVAVVV+SMKRDEETDKVHIRATIMVERDSQK
 Sbjct: 181 ITDHPERFLVSEMVREKVLHLLTREETPHSVAVVVSMSKRDEETDKVHIRATIMVERDSQK 240

Query: 256 GIIIGKGAMLKKIGKMARRDIELMLGDKVLETWVKVKNWRDKKLDLADFGYNEKEY 314
 GIIIGK GAMLKKIG MARRDIELMLGDKV+LETWVKVKNWRDKKLDLADFGYNE+EY
 Sbjct: 241 GIIIGKGAMLKKIGSMARRDIELMLGDKVLETWVKVKNWRDKKLDLADFGYNEREY 299

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6757> which encodes the amino acid sequence <SEQ ID 6758>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

-2466-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 295/297 (99%), Positives = 296/297 (99%)

10 Query: 18 FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTRNKIMGIYTTETEQIVFIDTPG 77
 FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTRNKIMGIYTTETEQIVFIDTPG
 Sbjct: 2 FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTRNKIMGIYTTETEQIVFIDTPG 61

15 Query: 78 IHKPKTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDDMIERLKAAKIPVILVINK 137
 IHKPKTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDDMIERLKAAKIPVILVINK
 Sbjct: 62 IHKPKTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDDMIERLKAAKIPVILVINK 121

20 Query: 138 IDKVHDPQLLEQIDDFRSQMDFKFVVPISALQGNNVPTLIKLLTDNLEEGFQYFPEDQIT 197
 IDKVHDPQLLEQIDDF SQMDFKFVVPISAL+GNNVPTLIKLLTDNLEEGFQYFPEDQIT
 Sbjct: 122 IDKVHDPQLLEQIDDFHSQMDFKFVVPISALEGNNVPTLIKLLTDNLEEGFQYFPEDQIT 181

25 Query: 198 DHPERFLVSEMVRKVLHLTQQEVPHSVAVVVE SMKRD EETDKVHIRATIMVERDSQKGI 257
 DHPERFLVSEMVRKVLHLTQQEVPHSVAVVVE SMKRD EETDKVHIRATIMVERDSQKGI
 Sbjct: 182 DHPERFLVSEMVRKVLHLTQQEVPHSVAVVVE SMKRD EETDKVHIRATIMVERDSQKGI 241

 Query: 258 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKNWRDKKLDLADFGYNEKEY 314
 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKNWRDKKLDLADFGYNEKEY
 Sbjct: 242 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKNWRDKKLDLADFGYNEKEY 298

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2188

A DNA sequence (GBSx2305) was identified in *S.agalactiae* <SEQ ID 6759> which encodes the amino acid sequence <SEQ ID 6760>. Analysis of this protein sequence reveals the following:

35 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2679(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2189

A DNA sequence (GBSx2306) was identified in *S.agalactiae* <SEQ ID 6761> which encodes the amino acid sequence <SEQ ID 6762>. Analysis of this protein sequence reveals the following:

50 Possible site: 21
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

 bacterial outside --- Certainty=0.3000(Affirmative) < succ>

-2467-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAA16793 GB:D90900 hypothetical protein [Synechocystis sp.]
 Identities = 36/119 (30%), Positives = 57/119 (47%), Gaps = 15/119 (12%)
 Query: 390 TSDYEKAKVIHDHLVNNITYATEELATTRETASGISIHAEALYKDKRGVCQAFVAMFKD 449
 ++D+E+A++ + + N Y +A TR I PE + +C ++ ++
 10 Sbjct: 153 SNDWEERLAYSWITQNIAYDVP-MAETRN---IDDLRPETVLARGETICSGYSNLYQA 207
 Query: 450 MAATAGLSVWYVTGQAGGG-----NHAWNIVTINGVKYYVDTTWDNNIKSNKYF 498
 +A GL V + G A GG NHAWN V I+G Y +DTTW I S+ F
 15 Sbjct: 208 LAKELGLDVVIIIEGFAKGGDVIVGDDPDVNHAWNNGVKIDGQWYLLDTTWGAGIVSDGKF 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6763> which encodes the amino acid sequence <SEQ ID 6764>. Analysis of this protein sequence reveals the following:

Possible site: 23
 20 >>> May be a lipoprotein
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 41/181 (22%), Positives = 79/181 (42%), Gaps = 17/181 (9%)
 30 Query: 355 ITITYTLKGDVGLHKEYKQFVDSFVKENITNKNITSDYEKAKVIHDHLVNNITYATE-- 412
 + +T+ + D +++++ Q + + + N +K+ YE+ K ++ ++ + Y +
 Sbjct: 124 VFVTFPIPEDAKNIYQDL-QAIGNDIVANTPSKD---RYEQVKYFYEVIIRD TDYNKKAF 179
 35 Query: 413 ELATTRETASGISIHAEALYKDKRGVCQAFVAMFKDMAATAGLSVWYVTGQAGCGN--- 469
 E + A S +++++ D VC +A F+ + AG+ V Y+ G
 Sbjct: 180 EAYQSGSQAQVASNQDIKSVFIDHLSVCNGYAQAQFLCQKAGIPVAYIRGTGTSQQPQQ 239
 40 Query: 470 ---HAWNIVTINGVKYYVDTTW-----DNNIKSNKYFLVGKTIMDADHLLDSQYNALAKDI 522
 HAWN V IN Y VD TW DN++ K + + + L + + +KDI
 Sbjct: 240 SFAHAWNVAQINNTYYGVDVTWGDVFDNHLSHQKQGTINYSFLCLPDYLMALSHQPSKDI 300

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

45 Example 2190

A DNA sequence (GBSx2307) was identified in *S.agalactiae* <SEQ ID 6765> which encodes the amino acid sequence <SEQ ID 6766>. This protein is predicted to be rgg protein. Analysis of this protein sequence reveals the following:

Possible site: 29
 50 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.16 Transmembrane 187 - 203 (187 - 203)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
 55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-2468-

A related GBS nucleic acid sequence <SEQ ID 10015> which encodes amino acid sequence <SEQ ID 10016> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAA26968 GB:M89776 rgg [Streptococcus gordonii]
    Identities = 71/273 (26%), Positives = 140/273 (51%), Gaps = 16/273 (5%)

    Query: 8  KELGKTLRRLRKGGKVSISSLADEHLSKSQISRFERGESEITCSRLNLNLDKLNITIDEF 67
              K  GK L+ +R+ K +S+ +A +S +Q+SR+ERG S +T + L +++++ EF
    Sbjet: 5  KSSGKILKTIRESKNMSLKEVAAGDISVAQLSRYERGISSLTVDSEFYSCLRNMSVSLAEF 64

10  Query: 68  VSI-HSKAHTHFFILLNRVRKYCAEKNVTKLVALL-----EDHNHKDYEKIMIK 115
              + H+ +L ++ + E N+ KL ++L E N+K I+I+
    Sbjet: 65  QYVYHNYREADDVVLSQKLSEAORENNIVKLESILAGSEAMAQEFPEKKNYK-LNTIVIR 123

15  Query: 116 ALIFSIDQSIENPQEEARLTLDYLFVEQWGYVEIILLGNC SRLINYNLFLLTKEMVNS 175
              A + S + + ++ ++ LTDYLF+VE+WG YE+ L N L+ TL EM+N
    Sbjet: 124 ATLTSCNPDYQVSKGDIEFLTDYLFVSEWGRYELWLFTNSVNLTLTLETFASEMINR 183

20  Query: 176 FAYSEQNKTINKILVTQLAINCLIIISIDHSYFEHSHYLIDKVRSLQDEVNFYEKTVFLYV 235
              + N+ + ++ +N+ I++++ + + ++ + + E + Y++ + Y
    Sbjet: 184 TQFYNNLPENRRRIIKMLLNVSACIENNHLQVAMKFLNYIDNTKIPETDLYDRVLIKYH 243

    Query: 236 TGYHHLKLGDTSSGKEDMRKALQIFKYLGEDSF 268
              Y K+G+ + + D+ + L F+YL DSF
25  Sbjet: 244 KALYSYKVCNPHA-RHDIEQCLSTFEYL--DSF 273
  
```

There is also homology to SEQ ID 628.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2191

A DNA sequence (GBSx2308) was identified in *S.galactiae* <SEQ ID 6767> which encodes the amino acid sequence <SEQ ID 6768>. Analysis of this protein sequence reveals the following:

```

    Possible site: 36
    >>> Seems to have no N-terminal signal sequence

35  ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3234(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40
  
```

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:BAA05066 GB:D26071 formamidopyrimidine-DNA glycosylase
      [Streptococcus mutans]
    Identities = 182/271 (67%), Positives = 217/271 (79%)

45  Query: 1  MPPELPEVETVRKGLERLVVNQEIASITIKVPMVKTDLNDFMISLPGKTIQQVLRGKYL 60
              MPPELPEVETVR+GLE L+V ++I S+ ++VPMVKVT + DF + + G+T + + RRGKYL
    Sbjet: 1  MPPELPEVETVRRGLEHLIVGKKIVSEVRVPMVKVTGVEDFQLDILGQTFESIGRRGKYL 60

50  Query: 61  LFDGEMVMVSHLRMEGKYLLFPNKPVDNKHFFHLYFKLTNGSTLVYQDVRKFGTFELVRK 120
              L + ++SHLRMEGKYLLF ++VPDNKHFFH+F L GSTLVYQDVRKFGTFEL+ K
    Sbjet: 61  LLNLNRQTIISHLRMEGKYLLFEDEVDPDNKHFFHFFGLDGGSTLVYQDVRKFGTFELLPK 120

    Query: 121 SSLKDYFTQKKLGPEPTADTFQEPFSGKGLANSKKPIKPLLLDQRLVAGLGNIYVDEVILW 180
              S ++ YF QKK+GPEP A F+ +PF +GLA S K IK LLLDQ LVAGLGNIYVDEVILW
55  Sbjet: 121 SQVEAYFVQKKIGPEPNKDFKLPFEEGLAKSHKVIKTLLLDQHLVAGLGNIYVDEVILW 180

    Query: 181 AAKIHPQRLANQLTESETSLHKEIIRILTLGIEKGGSTIRTYKNALGEDGTMQKYLQVY 240
  
```

-2469-

AAK+ P+RLA+QL SE +H E IRIL L IEKGGSTIR+YKN+LGEDG+MQ LQVY
 Sbjct: 181 AAKVDPERLASQLKTSEIKRIHDETIRILQLAIEKGGSTIRSYKNSLGEDGSMQDCLQVY 240

Query: 241 GKTGQPCPRCGCLIKKIKVGGRGTHYCPRCQ 271
 5 GKT QPC RC I+KIKVGGRGTH+CP CQ
 Sbjct: 241 GKTDQPCARCATPIEKIKVGGRGTHFCPCSCQ 271

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6769> which encodes the amino acid sequence <SEQ ID 6770>. Analysis of this protein sequence reveals the following:

10 Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.2068(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 190/271 (70%), Positives = 229/271 (84%)

20 Query: 1 MPPELPEVETVRKGLERLVVNQEIASITIKVPKMVKTDLNDFMISLPGKTIQQVLRGKYL 60
 MPPELPEVETVR+GLE LV+ QEI ++T+KVPKMVKTDL F ++LPG+ IQ V RRGKYL
 Sbjct: 1 MPPELPEVETVRRGLETLVLGQEIVAVTLKVPKMVKTDLETFTALTLPQIIQSVGRRGKYL 60

25 Query: 61 LFDGEMVMVSHLRMEGKYLLFPNKVPDNKHFFHLYFKLTNGSTLVYQDVRKFGTFELVRK 120
 L D G++V+VSHLRMEGKYLLFP++VPDNKHFFH++F+L NGSTLVYQDVRKFGTF+L+ K
 Sbjct: 61 LIDLQQLVLVSHLRMEGKYLLFPDEVPDNKHFFHVFELKNGSTLVYQDVRKFGTFDLIAK 120

30 Query: 121 SSLKDYFTQKKLGPEPTADTFQFEPFSKGLANSKKPIKPLLLDQRLVAGLGNIYVDEVLW 180
 S L +F ++KLGPPEP +TF+ + F L +S+KPIKP LLDQ LVAGLGNIYVDEVLW
 Sbjct: 121 SQLSAFFAKRKLGPPEPKKETFGLKTFEAALLSSQKPIKPHLLDQTLVAGLGNIYVDEVLW 180

Query: 181 AAKIHQRIANQITESETSLHKEIIRILTLGIEKGGSTIRTYKNALGEDGTMQKYLQVY 240
 AAK+HF+ +++L ++E LH E IRIL LGIEKGGST+RTY+NALG DGTMQ YLQVY
 35 Sbjct: 181 AAKVHPETASSRLNKAIEIKRLHDETIRILALGIEKGGSTVRYRNALGADGTMQDYLQVY 240

Query: 241 GKTGQPCPRCGCLIKKIKVGGRGTHYCPRCQ 271
 G+TG+PCPRCG I K+KVGGRGTH CP+CQ
 40 Sbjct: 241 GQTGKPCPRCGQAIVKLKVGGRGTHICPKCQ 271

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2192

45 A DNA sequence (GBSx2309) was identified in *S.agalactiae* <SEQ ID 6771> which encodes the amino acid sequence <SEQ ID 6772>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0797(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 10013> which encodes amino acid sequence <SEQ ID 10014> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00353 GB:AF008220 YtaG [Bacillus subtilis]

-2470-

Identities = 80/189 (42%), Positives = 113/189 (59%), Gaps = 1/189 (0%)

Query: 8 MTKIIGLTGGIASGKSTVTKIIRRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGP EILD 67
 MT +IGLTGGIASGKSTV ++ E G VIDAD + + KG Y+ +++ G +IL
 5 Sbjet: 1 MTLVIGLTGGIASGKSTVANMLIEKGITVIDADIIAKQAVEKGMPAYRQIIDEFGEDILL 60

Query: 68 ADGELDRPKLSQMIFANPDNMKTSARLQNSIIRQELACORDQLKQT EEIF-FMDIPLLE 126
 ++G++DR KL ++F N + + +RQE+ +RD+ E F +DIPL E
 10 Sbjet: 61 SNGDIDRKKLGALVFTNEQKRLALNAIVHPAVRQEMLNRRDEAVANREAFVVLDIPLLE 120

Query: 127 EKYIKWFDEIWLVFVDKEKQLQRLMARNNYSREEAELRLSHQMPLTDKKS FASLIIDNNG 186
 K D+I +V V KE QL+RLM RN + EEA R+ QMPL +K + A +IDN+G
 Sbjet: 121 SKLESIVDKIIVSVTKELQLERLMKRNQLTEEEAVSRIRSQMPLEBK TARADQVIDNSG 180

Query: 187 DLITLKEQI 195
 L K Q+
 15 Sbjet: 181 TLEETKRQL 189

A related sequence was also identified in GAS <SEQ ID 9111> which encodes the amino acid sequence
 20 <SEQ ID 9112>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty= 0.101(Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 118/191 (61%), Positives = 153/191 (79%)
 30 Query: 9 TKIIGLTGGIASGKSTVTKIIRRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGP EILDA 68
 T IIG+TGGIASGKSTV K+IR++G++VIDADQVVH LQ KGG+LY+AL E G +IL A
 35 Sbjet: 9 TMIIGITGGIASGKSTVVKIRKAGYQVIDADQVVHDLQEKGGRLYEALREAFGNQILKA 68

Query: 69 DGELDRPKLSQMIFANPDNMKTSARLQNSIIRQELACORDQLKQT EEIFFMDIPLLEEK 128
 DGELDR KLS+M+F+NPDNM TS+ +QN II++ELA +RD L Q++ IFFMDIPLLE+E
 Sbjet: 69 DGELDRTKLSEMLFSNPDNMATSSAIONQIIKEELAAKRDHLAQSQAIFFMDIPLLELG 128

Query: 129 YIKWFDEIWLVFVDKEKQLQRLMARNNYSREEAELRLSHQMPLTDKKS FASLIIDNNGDL 188
 Y WFD IWL+VD + QLQRLMARN + +A R++ Q+P+ +KK +ASL+IDN+GD+
 40 Sbjet: 129 YQDWFDIWLIVYVDAQTLQRLMARNRLDKGKARQRIASQLPIEEKPYASLVIDNSGDI 188

Query: 189 ITLKEQILDAL 199
 L +Q+ AL
 45 Sbjet: 189 AALIKQVQSAL 199

A related GBS gene <SEQ ID 8993> and protein <SEQ ID 8994> were also identified. Analysis of this
 protein sequence reveals a signal peptide at residues 1-16.

50 The protein has homology with the following sequences in the databases:

42.2/60.6% over 189aa

OMNI|NT01BS3382| Insert characterized

ORF02237(319 - 885 of 1206)
 55 OMNI|NT01BS3382(3 - 192 of 200) ()
 %Match = 17.0
 %Identity = 42.1 %Similarity = 60.5
 Matches = 80 Mismatches = 74 Conservative Sub.s = 35

78 108 138 168 198 228 258 288
 KNSPTAFG*SIDRI*NKLIITQGNYSHFNFRRKRRLHD*NI*ECSWRGRYDAKVFTGLW*NWATVSKVWLFN*EDKSRRE

5
RDALLPSVSM³¹⁸LMTKIIGLTGGIASGKSTVTTKI³⁴⁸IRESGFKVIDADQVVHKLOAKGGKL³⁷⁸YQALLEWLGP⁴⁰⁸EILDADGELDRPK⁴³⁸
| : | : ||||| : : | | ||| : : || | : : : : | : | : : ||| |
VDLLTLVLGLTGGIASGKSTVANMLIEKGITVIDADI⁴⁶⁸IAKQAVEKGMPAYRQII⁴⁹⁸IDFEGEDILLSNGDIDRKK⁵²⁸

10
LSQMIFANPDN⁵⁵⁸MKTSARLONSII⁵⁸⁸RQELACORDQLKQT⁶¹⁸EIFF-MDIPLLIBEKYIKWFDEIWL⁶⁴⁸VFDKEKQLQR⁶⁷⁵LMARNN⁷⁰⁵
| : : | | : : : ||| : : || : | | : |||| | | | : | : | | | | : ||| |
LGALVFTNEQKR⁷³⁵LALNAIVHPAVRQEMLNRREAVANREA⁷⁶⁵FVVDIPLLFESKLES⁷⁹⁵LVDKTIIVSV⁸²⁵TKEQLERIMKRNQ⁸⁵⁵
| : : | | : : : ||| : : || : | | : |||| | | | : | : | | | | : ||| |
LGTAAVFTNEQKR⁸⁸⁵LALNAIVHPAVRQEMLNRREAVANREA⁹¹⁵FVVDIPLLFESKLES⁹⁴⁵LVDKTIIVSV⁹⁷⁵TKEQLERIMKRNQ¹⁰⁰⁵

15
YSREEAELRLSHQ⁷⁹⁵MP⁸²⁵LTDDKKSFA⁸⁵⁵SLLIIDNNGDLITLKEQILDALQRL*N⁸⁸⁵*MDNVFIH⁹¹⁵PLSLH*F*KTC⁹⁴⁵D*TTVIVQ*Y⁹⁷⁵
| : ||| | : |||| : : | : ||| : | | | : : :
LTEEEAVSRIRS¹⁰⁰⁵QMPLEEK¹⁰³⁵TARADQVIDNSGTLEETKRQLDEIMNSWA¹⁰⁶⁵

Score = 218 bits (550), Expect = 4e-59
Identities = 104/175 (59%), Positives = 138/175 (78%)

25	Query: 25	VVKVIRKAGYQVIDADQVVHDLQEKGGRLYEALREAFGNQILKADGELDRITKLSMLFSN 84
		V K+IR++G++VIDADQVVH LQ KGG+LY+AL E G +IL ADGELDR KLS+M+F+N
	Sbjct: 20	VTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGP EILDADGELDRPKLSQMIFAN 79
30	Query: 85	PDNMATSSAIONQIIKEELAARKDHLAQSQAIFMDIPLLMLGYQDWFDAIWL VVYVDAQ 144
		PDNM TS+ +QN II++ELA +RD L Q++ IFMDIPLL+E Y WFD IWL V+VD +
	Sbjct: 80	PDNMKTSARLQNSIIRQELACQDQLKQTEEIFMDIPLLIEEKYIKWFDEIWL VFDKE 139
35	Query: 145	TQLQRLMARNRLDKGKARQRIASQLPIEEKPYASLVIDNSGDI AALIKVQSAL 199
		QLQRLMARN + +A R++ QP+ +KK +ASL+IDN+GD+ L +Q+ AL
	Sbjct: 140	KOLORLMARNNYSREEEAE LRLSHOMPLTDKKSFASLIIDNNGDLITLKEOILDAL 194

40 The purified GST fusion product was used to immunise mice and the resulting antiserum was used for FACS (Figure 278). This confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2193

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
```

```
50      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4073(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

>GP:BAA30330 GB:AP000005 253aa long hypothetical ATP-binding

-2472-

transport protein [Pyrococcus horikoshii]
 Identities = 78/240 (32%), Positives = 130/240 (53%), Gaps = 13/240 (5%)

Query: 3 LVIRDIRKRFQETEVLRGASYRFYSGKITGVLRNGAGKTTLFNIIYGDIAADNGTICLL 62
 +++ ++RK+F EVL+G ++ G+I G+LG NG+GK+T IL G + G + +
 5 Sbjct: 2 IIVENLRKKFGSKFVLKGINFTVNDGEIYGLLPNGSGKSTIMRILSGIITDFEGKVMVA 61

Query: 63 -KDNHEYPLTDKDI-GIVYSENYLPEFLTGYEFVKFYMDLH--PSDDL-MTIDDYLDLDFME 117
 D P+ K+I G V L E LT EF F + P D L + +D
 10 Sbjct: 62 GVDVSRDPMKVKEIVGYVPETPALYESLTPAEFFSFIGGVRRIPQDILEERVKRLVDAFG 121

Query: 118 IGQTERHRIIKGYSDGMKSKLSLCLMISKPKVILLDEPLTAVDVVSSIAIKRLLELSE 177
 IG+ +++I S G K K+SLI ++ P+V++LDE + +D S+ + LL E E
 15 Sbjct: 122 IGK-YMNQLIGTILSFGTKQKISLISALLHDPQVLILDEAMNGLDPKSARIFRELLFEFKE 180

Query: 178 D-HIIILSTHIMALAEDLCDIVAVLDKGL---QTLDIDR---KHEQFEERLLQVLKGDE 230
 + I+ STHI+ALAE +CD + ++ +G++ T+D R + E+ E+ L++ + E
 Sbjct: 181 EGKSIVFSTHILALAEVMCDRIGIYEGRIVAEGTIDELREIAREEKLIEDIFLKLQAKE 240

20 There is also homology to SEQ ID 2876.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2194

A DNA sequence (GBSx2311) was identified in *S.agalactiae* <SEQ ID 6777> which encodes the amino
 25 acid sequence <SEQ ID 6778>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.6138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2195

A DNA sequence (GBSx2312) was identified in *S.agalactiae* <SEQ ID 6779> which encodes the amino
 40 acid sequence <SEQ ID 6780>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -15.34	Transmembrane	526 - 542 (511 - 546)
INTEGRAL	Likelihood = -9.61	Transmembrane	340 - 356 (335 - 359)
45 INTEGRAL	Likelihood = -8.17	Transmembrane	455 - 471 (451 - 476)
INTEGRAL	Likelihood = -8.01	Transmembrane	97 - 113 (95 - 121)
INTEGRAL	Likelihood = -8.01	Transmembrane	216 - 232 (207 - 236)
INTEGRAL	Likelihood = -3.40	Transmembrane	50 - 66 (46 - 67)
50 INTEGRAL	Likelihood = -1.33	Transmembrane	178 - 194 (178 - 194)

----- Final Results -----
 bacterial membrane --- Certainty=0.7135(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2473-

A related GBS nucleic acid sequence <SEQ ID 10011> which encodes amino acid sequence <SEQ ID 10012> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is
5 homology to SEQ ID 376.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2196

10 A DNA sequence (GBSx2314) was identified in *S.galactiae* <SEQ ID 6781> which encodes the amino acid sequence <SEQ ID 6782>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence
15  INTEGRAL    Likelihood = -8.17    Transmembrane  140 - 156 ( 134 - 160)
    INTEGRAL    Likelihood = -6.64    Transmembrane  255 - 271 ( 253 - 274)
    INTEGRAL    Likelihood = -5.79    Transmembrane  345 - 361 ( 343 - 363)
    INTEGRAL    Likelihood = -3.29    Transmembrane  184 - 200 ( 183 - 202)
    INTEGRAL    Likelihood = -2.34    Transmembrane   66 -  82 (  65 -  83)
    INTEGRAL    Likelihood = -1.65    Transmembrane  221 - 237 ( 221 - 239)
20  INTEGRAL    Likelihood = -0.00    Transmembrane  121 - 137 ( 121 - 137)

----- Final Results -----
        bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9401> which encodes amino acid sequence <SEQ ID 9402> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:CAA07482 GB:AJ007367 multi-drug resistance efflux pump
    [Streptococcus pneumoniae]
    Identities = 213/372 (57%), Positives = 295/372 (79%)

35  Query: 1  MPFMVLYVEQLGAPSNKVEWYAGLSVLSALSSALVAPLWGRLADKYGRKPMVMVRAGLMM 60
    +PFM ++VE LG S +V +YAGL++S+SA+S+AL +P+WG LADKYGRKPM+RAGL M
    Sbjct: 28  VPFMPIFVENLGVGSQQVAFYAGLAISVSAISAAFLSPIWGLADKYGRKPMIRAGLAM 87

    Query: 61  TFTMGGLAFIHSVTGLLILRIILNGIFAGYVPNSTALIASQAPQEESGYALGTLATGVTGG 120
    T TMGGLAF+ ++ L+ LR+LNG+FAG+VPN+TALIASQ P+E+SG ALGTL+TGV G
40  Sbjct: 88  TITMGGLAFVPNIYWLIFLRLNGVFAGFVPNATALIASQVPKEKSGSALGTLSTGVVAG 147

    Query: 121 MLIGPLLGLLAEWFGIREVFLLVGTILLISTIMTIFMVKEDFKPISNEETMPTEVFVK 180
    L GP +GG +AE FGIR VFLLVG+ L ++ ++TI +KEDF+P++ E+ +PT E+F S
45  Sbjct: 148 TLTGPFIFGFIAEFGIRTVPFLVGSFLLAAILTICFIKEDFQVAKEKAIPKELFTS 207

    Query: 181 VKSLQILIGLFVTSMIQISAQSIAPILTLYIRHLGQTENLMFVSGLIVSGMGFSSILSS 240
    VK +L+ LF+TS +IQ SAQSI PIL LY+R LQQTENL+FVSGLIVS MGFSS++S+
    Sbjct: 208 VKYPYLLLLNLFVTSFVIQFSAQSIGPILALYVRDLGQTENLFLVSGLIVSSMGFSSMMSA 267

50  Query: 241 PKLGRIGDRIGNHRLLLLALLYSFLMYVLCSLAQTSLQLGVIRFLYGFGTGALMPSINSI 300
    +G++GD++GNHRL++A YS ++Y+LC+ A + LQLG+ RFL+G GTGAL+P +N++
    Sbjct: 268 GVMGKLGDKVGNNHRLLVVAQFYSVIYLLCANASSPLQLGLYRFLRGLGTGALIPGVNAL 327

55  Query: 301 LTKIAPRQGLSRIFSYNQMFSNLGQVLGPFVGSVAVSIHLGFRWVFFVTSFIVLANFVWCF 360
    L+K+ P+ G+SR+F++NQ+F LG V+GP GSAV+ G+ VF+ TS V + ++

```

-2474-

Sbjct: 328 LSKMTPKAGISRVAFAFNQVFFYLGGVVGPMAGSAVAGQFGYHAVFYATSLCVAFAFSCLEFNL 387

Query: 361 INFRKYIRVKEI 372
I FR ++VKEI

5 Sbjct: 388 IQFRTLLKVKEI 399

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6783> which encodes the amino acid sequence <SEQ ID 6784>. Analysis of this protein sequence reveals the following:

Possible site: 58
10 >>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -10.14 Transmembrane 165 - 181 (150 - 185)
INTEGRAL Likelihood = -7.43 Transmembrane 371 - 387 (367 - 391)
INTEGRAL Likelihood = -3.88 Transmembrane 90 - 106 (86 - 109)
15 INTEGRAL Likelihood = -3.35 Transmembrane 145 - 161 (143 - 162)
INTEGRAL Likelihood = -1.70 Transmembrane 279 - 295 (279 - 297)
INTEGRAL Likelihood = -0.85 Transmembrane 209 - 225 (209 - 226)
INTEGRAL Likelihood = -0.27 Transmembrane 347 - 363 (347 - 363)
20 ----- Final Results -----
bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:CAA07482 GB:AJ007367 multi-drug resistance efflux pump
[Streptococcus pneumoniae]
Identities = 236/396 (59%), Positives = 309/396 (77%)
30 Query: 1 VNWRQNLKVAWLGNFFTGASFSLVMPFMALYVENLGTPTLVEYYAGLAVAVTALASALF 60
+NW+ NL++AW GNF TGAS SLV+PFM ++VENLG ++ V +YAGLA++V+A+++ALF
Sbjct: 4 INWKDNLRIAWFGNFLTGTASISLVVPMPIFVENLGVGSQQVAFYAGLAISVSAISAALF 63
Query: 61 APVWGKLADRYGRKPMMLRASFMVMTFTMGGLAIIPNVFWLLILRLLTGVSAGYVFNATAL 120
+P+WG LAD+YGRKPM+RA MT TMGGLA +PN++WL+ LRL L GV AG+VFNATAL
35 Sbjct: 64 SPIWGLADKYGRKPMIRAGLAMTITMGGLAFVFNWLIIFLRLINGVFAGFVFNATAL 123
Query: 121 IASQAPKEESGYALGTLATGVTAGALIGPLLGILAEELLGIRQVFLLVGVILFLCSLMTA 180
IASQ PKE+SG ALGTL+TGV AG L GP +GG +AEL GIR VFLLVG LFL +++T
40 Sbjct: 124 IASQVKEKSGSALGTLSTGVVAGTLTGPFIFGGFIAELFGIRTVFLLVGSFLFLAAILTI 183
Query: 181 VYVKEEFKPVRRFEMIPKVIKQVKSPQIMLGLFVTSMIQISAQSVAPILSLYIRHLG 240
++KE+F+PV + + IPTK + VK P ++L LF+TS +IQ SAQS+ PIL+LY+R LG
Sbjct: 184 CFIKEDFQPVAKKAIPKELFTSVKYPYLLNLFLTFSVIOFSAQSIGPILALYVRDLG 243
45 Query: 241 QTHNLMFTSGLVVSAMGFSSSLFSSSYLGKIGDRFGNHRLLLAALCYSFIMYFSSALAQT 300
QT NL+P SGL+VS+MGFSS+ S+ +GKLG+ GNHRL+ A YS I+Y A A +
Sbjct: 244 QTENLLFVSGLIVSSMGFSSMMSAGVMGKLGDKVGNHRLLVVAQFYSVIITLLCANASSP 303
50 Query: 301 FQLGVLRFAFGFGVGMPSINSLLTKLTPKEGISRVFAYNQMFNSNLGQVIGPFIGSNVA 360
QLG+ RF +G G GAL+P +N+LL+K+TPK GISRVFA+NQ+F LG V+GP GS VA
Sbjct: 304 LQLGLYRFLFGLGTGALIPGVNALLSKMTPKAGISRVAFAFNQVFFYLGGVVGPMAGSAVA 363
Query: 361 VVLGYRSVFYVTSLVFVNLIWSLIIFRKYIKVKDI 396
GY +VFY TSL V + +++LI FR +KVK+I
55 Sbjct: 364 GQFGYHAVFYATSLCVAFAFSCLEFNL IQFRTLLKVKEI 399

An alignment of the GAS and GBS proteins is shown below.

Identities = 262/373 (70%), Positives = 314/373 (83%)
60 Query: 1 MPFMVLYVEQLGAPSNKVEWYAGLSVLSALSSALVAPLWGRLADKYGRKPMVVRAGLMM 60
MPFM LYVE LG P+ VE+YAGL+V+++AL+SAL AP+WG+LAD+YGRKPM+RA +M
Sbjct: 25 MPFMALYVENLGTPTLVEYYAGLAVAVTALASALFAPVWGKLADRYGRKPMMLRASFMV 84
Query: 61 TFTMGGLAFIHSVTGLLILRLINGIFAGYVFNSTALIASQAPQEEESGYALGTLATGVTGG 120

-2475-

TFTMGGLA I +V LLILR+L G+ AGYVPN+TALIASQAP+EESGYALGTLATGVT G
 Sbjct: 85 TFTMGGLAIPNVFWLLILRLTGVSAAGYVPNATALIASQAPKEESGYALGTLATGVTAG 144
 Query: 121 MLIGPLLGGLLAEWFGIREVFLLVGTILLISTLMTIFMVKEDFKPISNEETMPTTEVFKS 180
 5 LIGPLLGG+LAE GIR+VFLLVG IL + +LMT VKE+FKP+ E +PT + K
 Sbjct: 145 ALIGPLLGGILAEELLGIRQVFLLVGVILFLCSLMTAVYVKEEFKPVRRFEMIPTKVILKQ 204
 Query: 181 VKSLQILIGLFTVSMIIQISAQSIAPILTLIRHLGQTNLMFVSGLIVSGMGFSSILSS 240
 10 VKS QI++GLFVTSIIQISAQS+APIL+LYIRHLGQT NLMF SGL+VS MGFSS+ SS
 Sbjct: 205 VKSPQIMLGLFTVSMIIQISAQSVAPILSLYIRHLGQTNLMFTSGLVVSAMGFSSILSS 264
 Query: 241 PKLGRIGDRIGNHRLLLALLYSFLMYVLCSLAQTSLQLGVIRFLYGFGTGALMPSINSI 300
 15 LG++GDR GNHRLLL AL YSF+MY +LAQTS QLGV+RF YGFG GALMPSINS+
 Sbjct: 265 SYLGKLGDRFGNHRLLLAALCYSFIMYFSSALAQTSFQLGVLRFPAYGFGVGVGALMPSINSL 324
 Query: 301 LTKIAPRQGLSRIFSYNQMFNSNLQVLPFVGSVAVSIHLGFRVWFFVTSFIVLANFVWCF 360
 20 LTK+ P++G+SR+F+YNQMFNSNLQV+GPF+GS V++ LG+R VF+VTS IV N +W
 Sbjct: 325 LTKLTPKEGISRVFAYNQMFNSNLQVIGPFIGSNVAVVLGYRSVFVYVTSFIVFVNLWISL 384
 Query: 361 INFRKYIRVKEIV 373
 I FRKYI+VK+IV
 Sbjct: 385 IIFRKYIKVKDIV 397

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2197

A DNA sequence (GBSx2315) was identified in *S. galactiae* <SEQ ID 6785> which encodes the amino acid sequence <SEQ ID 6786>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2343(Affirmative) < succ>
 35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB69986 GB:U94356 glycerol kinase [Enterococcus faecalis]
 Identities = 156/186 (83%), Positives = 167/186 (88%), Gaps = 1/186 (0%)
 40 Query: 3 SEEKYIMAIHQGTSSRAIIFNKKGEKIASQKEFPQIFPQAGWVEHNANQIWNVSQSVI 62
 +BEKYIMAIHQGTSSRAIIF+KKG KI SSQKEF Q FP AGWVEHNAN+IWNVSQSVI
 Sbjct: 2 ABEKYIMAIHQGTSSRAIIFDKKGNKIGSSQKEFTQYFPNAGWVEHNANEIWNVSQSVI 61
 45 Query: 63 AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122
 AG+ IES +KP I IGITNQRETTVVWDK TGLPIYNAIVWQSRQT PIADQLK++G+
 Sbjct: 62 AGSLIESGVKPTDIAGIGITNQRETTVVWDKATGLPIYNAIVWQSRQTTPADQLKEDGY 121
 50 Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPQAQERAEKGELLFGTIDTWLVWKLTDGLVHV 182
 + MIHEKTGL+IDAYFSATKVRWILDHV GAQERAE GEL+FGTIDTWLVWKL T G HV
 Sbjct: 122 SEMIHEKTGLIIDAYFSATKVRWILDHVEGAQERAEENGELMFGTIDTWLVWKLTDGLVHV 180
 Query: 183 TDYSNA 188
 TDYSNA
 55 Sbjct: 181 TDYSNA 186

There is also high homology to SEQ ID 2844:

Identities = 174/186 (93%), Positives = 182/186 (97%)
 60 Query: 3 SEEKYIMAIHQGTSSRAIIFNKKGEKIASQKEFPQIFPQAGWVEHNANQIWNVSQSVI 62
 S+EKYIMAIHQGTSSRAIIFN+KGEK++SSQKEFPQIFP AGWVEHNANQIWNVSQSVI

-2476-

Sbjct: 2 SQEKYIMAIQGTSSRAIIFNQKGEKVSSSQKEFPQIFPHAGWVEHNNANQIWNVSQSVI 61

Query: 63 AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122
 AGAFIESSIKP QIEAIGITNQRETTVVWDKKTG+PIYNAIVWQSRQTAPIA+QLKQ+GH

5 Sbjct: 62 AGAFIESSIKPSQIEAIGITNQRETTVVWDKKTGVPIYNAIVWQSRQTAPIAEQLKQDGH 121

Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPGAQERAKEGELLFGTIDTWLVWKLTDGLVHV 182
 T MIHEKTGLVIDAYFSATK+RWILDHVPGAQERAKEGELLFGTIDTWLVWKLTDG VHV

10 Sbjct: 122 TKMIHEKTGLVIDAYFSATKIRWILDHVPGAQERAKEGELLFGTIDTWLVWKLTDGAVHV 181

Query: 183 TDYSNA 188
 TDYSNA

Sbjct: 182 TDYSNA 187

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2198

A DNA sequence (GBSx2317) was identified in *S.galactiae* <SEQ ID 6787> which encodes the amino acid sequence <SEQ ID 6788>. This protein is predicted to be glycyl-tRNA synthetase beta chain (glyS).

20 Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2933(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAB14468 GB:Z99117 glycyl-tRNA synthetase (beta subunit)
 [Bacillus subtilis]
 Identities = 315/687 (45%), Positives = 447/687 (64%), Gaps = 21/687 (3%)

35 Query: 3 KDLLLELGLLEELPAYVVTPESEKQLGQKMKVLFEDHRLSFETVQIFSTPRRLAVRVKGLAD 62
 +DLLLE+GLEE+PA + S QLG K+ +L++ ++ V++F+TPRRLAV VK +A+

Sbjct: 4 QDLLLEIGLEEMPARFLNESMVQLGDKLTGWLKEKNITHGEVKLENTPRRLAVFVKDVAE 63

Query: 63 QQTDLTDFKGPSKKIALDAEGNFSKAAQGFVRGKGLSVDDIEFREVKGEEYVYVTKHET 122
 +Q D+ E+ KGP+KKIALDA+GN++KAA GF +G+G +V+D+ +EVKG EYV+V K +

40 Sbjct: 64 KQDDIKKEAKGPAKKIALDADGNWTKAAIGFSKGGQGANVEDLYIKVKGIEYVVFQKQFA 123

Query: 123 GKSAIDVLASVTEVLTELTFPVNMHWANNSFEYIRPVHTLVLLDDQALELDFLDIHSGR 182
 G+ +L ++ ++T L FP NM W N YIRP+ +V L + ++ SGR

45 Sbjct: 124 GQETKSLLEPELSGLITSLHFPKNMRWGNEDLRYIRPIKWIVALFGQDVIPFSITNVESGR 183

Query: 183 ISRGHRFLGSDTEISSASSYEDDLRQQFVIADAKERQQMIVNQIHAIIEKKNISVEIDED 242
 ++GHRFLG + I S S+YE+ L+ Q VIAD R+QMI +Q+ + + N S+ +DED

50 Sbjct: 184 TTQGHRFLGHEVSIESPSAYEEQLKGQHVIAIDPSVRKQMIQSQLETMAAENNWSIPVDED 243

Query: 243 LLNEVLNLVEYPTAFLGSEFDEKYLDPPEEVLVTSMKNHQRYFVVRDRDGKLLPNFISVRN 302
 LL+EV +LVEYPTA GSF+ ++L +PEEVLVT+MK HQRYF V+D++G LLP+FI+VRN

55 Sbjct: 244 LLEDEVNHLVEYPTALYGSFESEFLSIPEEVLVTMKEHQRYFPVKDKNGDLLPHFITVRN 303

Query: 303 GNAEHIENVIKNEKVLVARLEDGEFFWQEDQKLNIAIDLVEKLKQVTFHEKIGSLYEHMD 362
 GN+ IENV +GNEKVL ARL D FF++EDQKLN I V+KL+ + FHE++GSL + +

60 Sbjct: 304 GNSHAIENVARGNEKVLRLRLSDASFFYKEDQKLNIDANVKLENIVFHEELGSLADKVR 363

Query: 363 RVKVISQYLAEKADLSDEEKLAVLRAASIYKFDLLTGMVDEFDELQGIMGEKYALLAGEQ 422
 RV I++ LA + ++ V RAA I KFDL+T M+ EF ELQGIMGEKYA + GE

Sbjct: 364 RVTSTIAEKLAVRLQADEDTLKHVKRAAEISKFDLVTHMIYEFPELQGIMGEKYARMLGED 423

-2477-

Query: 423 PAVAAAIREHYMPTSADGELPETRVGAILALADKFDTLSSFFSVGLIPSGSNDPYALRRA 482
 AVAAA+ EHYMP SA GE P T GA++A+ADK DT+ SFFS+G+IP+GS DPY L R
 Sbjct: 424 EAVAAAVNEHYMPSAGGETPSTFTGAVVAMADKLDTIASFFSIGVIPTGSQDPYGLPRQ 483

5 Query: 483 TQGIIVRILEAFGWDIPLDELVTNLVGLSFASLDYANQKEVMAFISARIEKMIGS-KVPKD 541
 GIV IL W I +EL+T F D N E++ F + R++ ++ + ++ D
 Sbjct: 484 ASGIVAILLDRNWGISFEELT-----FVQTDKEN--ELLDFFTQRLKYVLNAEQIRHD 535

10 Query: 542 IREAVLESPTYIVSLILEASQALVQKSKDAQYKVSVESLSRAFNLAEKVTHSVLVDSSLF 601
 + +AVLES L +Q L QK +K + E+L R ++++K + LF
 Sbjct: 536 VIDAVLESSELEPYSALHKAQVLEQKLGAPGFKETAEALGRVISISKKGVGRGD-IQPDLF 594

15 Query: 602 ENNQEKALYQAILSLELTEDMHDNLDK-----LFALSPIINDFFDNTMVMTDDEKM 652
 EN E L+ A + + E++ +N K L AL I+ +FD+TMV+ D+E +
 Sbjct: 595 ENEYEAKLFDAYQTAK--ENLQENFSKKDYEAALASLAALKEPIDAYFDHTMVIADNESL 652

20 Query: 653 KQNRLAILNSLVAKARTVAAFNLLNTK 679
 K NRLA + SL + ++ A N L K
 Sbjct: 653 KANRLAQMVSLADEIKSFANMNALIVK 679

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2835> which encodes the amino acid sequence <SEQ ID 2836>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -0.96 Transmembrane 450 - 466 (450 - 466)

----- Final Results -----
 bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 505/679 (74%), Positives = 578/679 (84%)

35 Query: 1 MTKDLLLELGLLELPAYVVTPEKQLGQKMKFLEDHRLSFETVQIFSTPRRLAVRVKGL 60
 M+K+LL+ELGLEELPAYVVTPEKQLG+++ FL ++RLSFE +Q FSTPRRLAVRV GL
 Sbjct: 1 MSKNLLLELGLLELPAYVVTPEKQLGERLATFTENRLSFEDIQTFSTPRRLAVRVSGL 60

40 Query: 61 ADQQTDLTDFKGPSKKIALDAEGNFSKAAQGFVRGKGLSVDDIEFREVKGEEYVYVTKH 120
 ADQQTDLTDFKGP+KKIALDA+GNFSKAAQGFVRGKGL+ D IEFREVKGEEYVYVTKH
 Sbjct: 61 ADQQTDLTDFKGPAAKIALDADGNFSKAAQGFVRGKGLTTDAIEFREVKGEEYVYVTKH 120

45 Query: 121 ETGKSAIDVLASVTEVLTELTFPVNMHWANNSFEYIRPVHTLVLLDDQALELDFLDIHS 180
 E GK A +VL VTEVL+ +TFPV+MHWANNSFEYIRPVHTL VLL+D+ALELDFLDIHS
 Sbjct: 121 EAGKPAKEVLLGVTEVLSAMTFPVSMMHWANNSFEYIRPVHTLVLLNDEALELDFLDIHS 180

50 Query: 181 GRISRGHRFLGSDTEISSASSYEDDLRQQFVIADAKERQQMIVNQIHAIEEKNISVEID 240
 GR+SRGHRFLG++T I+SA SYE DLR QFVIADAKERQ+MIV QI +E ++ + V+ID
 Sbjct: 181 GRVSRGHRFLGTETTTTSADSYEADLRSQFVIADAKERQEMIVEQIKTLEVEQGVQVDID 240

55 Query: 241 EDLLNEVLNLVEYPTAFLGSFDEKYLDPPEEVLVTSMKNHQRYFVVRDRDGKLLPNFISV 300
 EDLLNEVLNLVE+PTAF+GSF+ KYLDVPEEVLVTSMKNHQRYFVVRD+ G L+PNF+SV
 Sbjct: 241 EDLLNEVLNLVEFPTAFMGSPFAKYLDVPEEVLVTSMKNHQRYFVVRDQAGHLMPNFVS 300

60 Query: 301 RNGNAEHIENVIKNEKVLVARLEDGEFFWQEDQKLNIALDLVEKLGQVTFHEKIGSLYEH 360
 RNGN + IENVIKNEKVLVARLEDGEFFW+EDQKL IADLV KL VTFHEKIGSL EH
 Sbjct: 301 RNGNDQAIENVIKNEKVLVARLEDGEFFWREDQKLQIADLVAKLTNVTFHEKIGSLAEH 360

65 Query: 361 MDRVKVISQYLAEKADLSDEEKLAVLRAASIYKFDLLTGMVDFDELQGIMGEKYALLAG 420
 MDR +VI+ LA++A+LS EE AV RAA IYKFDLLTGMV EFDELQGIMGEKYALLAG
 Sbjct: 361 MDRTRVIAASLAKEANLSAEEVTAVDRAAQIYKFDLLTGMVGEFDELQGIMGEKYALLAG 420

Query: 421 EQPAVAAAIREHYMPTSADGELPETRVGAILALADKFDTLSSFFSVGLIPSGSNDPYALR 480
 E AVA AIREHY+P +A G LPET+VGA+LALA K DTLSSFFSVGLIPSGSNDPYALR
 Sbjct: 421 EDAAVATAIREHYLPDAAGGALPETKVGAVLALAAKLDTLSSFFSVGLIPSGSNDPYALR 480

-2478-

Query: 481 RATQGIVRILEAFGWDIPLDELVTNLYGLSFASLDYANQKEVMAFISARIEKMIGSKVPK 540
 RATQGIVRIL+ FGW IP+D+LV +LY LSF SL YAN+ +VM FI AR++KM+G PK
 Sbjct: 481 RATQGIVRILDHFGWRIPMDKLVDLSYDLSFDSLTYANKADVMNFIRARVDKMMGKAAPK 540

Query: 541 DIREAVLESPTYIVSLILEASQALVQKSKDAQYKVSVESLSRAFNLAEKVTHSVLVDSSL 600
 DIREA+L S T++V +L A++ALV+ S YK +VESLSRAFNLAEK SV VD SL
 Sbjct: 541 DIREAILASSTFVVPPEMLAAAEALVKASHTENYKPAVESLSRAFNLAEKADASVQVDPFL 600

Query: 601 FENNQEKALYQAILSLLETTEDMHDNLDKLFALSPINDFFDNTMVMTDDEKMKQNLRLAIL 660
 FEN QE L+ AI L L L+++FALSP+INDFFDNTMVM D+ +K NRLAIL
 Sbjct: 601 FENEQENTLFAAIQGLTLAGSAAQQLQVLFALSPVINDFFDNTMVMAGDQALKNNRLAIL 660

Query: 661 NSLVAKARTVAAFNLNLTN 679
 + LV+KA+T+ AFN LNTK
 Sbjct: 661 SDLVSKAKTIVAFNQLNLTN 679

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2199

A DNA sequence (GBSx2318) was identified in *S.agalactiae* <SEQ ID 6789> which encodes the amino acid sequence <SEQ ID 6790>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2182(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD24436 GB:AF112858 NAD(P)H dehydrogenase [Bacillus
 stearothermophilus]
 Identities = 64/174 (36%), Positives = 98/174 (55%), Gaps = 6/174 (3%)

Query: 2 NTLIVNSHPDFSNPYSFTTILQEKFIELYNEHFENHQLSIILNLYDCVLPEITKEVLLSIW 61
 N L + +HP + S++ + + FI+ Y + P+H++ L+LY +PEI +V S W
 Sbjct: 3 NVLYITAHPH-DDTQSYSMAGVKAFFIDTYKQVHPDHEVIHLDLKEYIPEIDVDVF-SGW 60

Query: 62 SKQRKGL---ELTADEIVQAKISKDLLEQFKSHHRIVFVSPMHNYNVTARAKTYIDNIFI 118
 K R G EL+ +E + +L EQF S + VFV+PM N++ K YID + +
 Sbjct: 61 GKLRSGKSFEELSDEEKAKVGRMNEICEQFISADKYVFVTPMWNFSFPPVLKAYIDAVAV 120

Query: 119 AGETFKYTENGSVGLMTIDYRLIMLESAGSIYSKGQYSPYEFVHYLKAIFKDF 172
 AG+TFKYTE G VGL+TD + L +++ G YS+G + E YL I + F
 Sbjct: 121 AGKTFKYTEQGPVGLLTDK-KALHIQARGGFYSEGPAEMEMGHRYLSVIMQFF 173

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2200

A DNA sequence (GBSx2319) was identified in *S.agalactiae* <SEQ ID 6791> which encodes the amino acid sequence <SEQ ID 6792>. This protein is predicted to be glycyl-tRNA synthetase (glyQ). Analysis of this protein sequence reveals the following:

Possible site: 56

-2479-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1364 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9521> which encodes amino acid sequence <SEQ ID 9522> was also identified.

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05089 GB:AP001511 glycyl-tRNA synthetase (alpha subunit)
 [Bacillus halodurans]
 Identities = 222/287 (77%), Positives = 250/287 (86%)

15 Query: 6 LTFQEIIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPSRRPA 65
 + Q +ILTLQ++W+ Q C+L+QAYD EKGAGTMSPYT LR IGPEPWNAAYVEPSRRPA
 Sbjct: 1 MNVQIMILTLQEQYWSKQNCILLQAYDTEKGAGTMSPYTMLRTIGPEPWNAVAYVEPSRRPA 60

20 Query: 66 DGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLKSLELLGINPLEHDIRFVEDNWNENPST 125
 DGRYGENPNRLYQHHQFQV+MKPSP+NIQELYL SL LGINPLEHDIRFVEDNWNENPS
 Sbjct: 61 DGRYGENPNRLYQHHQFQVIMKPSPTNIQELYLDSLRALGINPLEHDIRFVEDNWNENPSL 120

25 Query: 126 GSAGLGWEVWLDGMEITQFTYFQQVGGGLQTGPVTSEVTYGLERLASYSIQEVDVSVYDIWA 185
 G AGLGWEVWLDGMEITQFTYFQQVGGGL+ PV++E+TYGLERLASYSIQ+ ++V+D+EW
 Sbjct: 121 GCAGLGWEVWLDGMEITQFTYFQQVGGLEANPVSAEITYGLERLASYSIQDKENVFDLEWV 180

30 Query: 186 PGVKYGEIFTQPEYEHKYSFEISDQVMLLENFEKFEREAKRALEEGLVHPAYDYVLKCS 245
 G YG+IFTQPEYEHKYS+FE+SD ML E F +E+EA RALEE LV PAYDYVLKCS
 Sbjct: 181 EGFTYGDIFTQPEYEHKYSKTFEVSDSAMLFELFSTYEKEADRALEENLVFPAYDYVLKCS 240

35 Query: 246 HTFNLLDARGAVSVTERAGYIARINLARVVAKTFAERKKLGFPLL 292
 HTFNLLDARGA+SVTER GYI R+RNLAR AK + ER+KLGFPP+L
 Sbjct: 241 HTFNLLDARGAISVTERTGYIGVRNLRKCAKKYEEEREKLGFPML 287

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6793> which encodes the amino acid sequence <SEQ ID 6794>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.2081 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 290/304 (95%), Positives = 294/304 (96%)

50 Query: 2 MSKKLTFQEIIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS 61
 MSKKLTFQEIIILTLQQ+WNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS
 Sbjct: 1 MSKKLTFQEIIILTLQQYWNQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS 60

55 Query: 62 RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLKSLELLGINPLEHDIRFVEDNWE 121
 RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYL SLE LGINPLEHDIRFVEDNWE
 Sbjct: 61 RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLASLEKLGINPLEHDIRFVEDNWE 120

60 Query: 122 NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGGLQTGPVTSEVTYGLERLASYSIQEVDVSVYD 181
 NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGGL T PVT+EVITYGLERLASYSIQEVDVSVYD
 Sbjct: 121 NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGGLATSPVTAEVITYGLERLASYSIQEVDVSVYD 180

60 Query: 182 IEWAPGVKYGEIFTQPEYEHKYSFEISDQVMLLENFEKFEREAKRALEEGLVHPAYDYV 241
 IEWAPGVKYGEIF QPEYEHKYSFEISDQ MLLENFEKFE+EA RALEEGLVHPAYDYV
 Sbjct: 181 IEWAPGVKYGEIFLQPEYEHKYSFEISDQMLLENFEKFEKASRALEEGLVHPAYDYV 240

-2480-

Query: 242 LKCSHTFNLLDARGAVSVTERAGYIARINLARVVAKTFVAERKKLGFFLLDEETRIKLL 301
 LKCSHTFNLLDARGAVSVTERAGYIARINLARVVAKTFVAERKKLGFFLLDE TR LL
 Sbjct: 241 LKCSHTFNLLDARGAVSVTERAGYIARINLARVVAKTFVAERKKLGFFLLDEATRAILL 300

Query: 302 AEED 305
 AE+D
 Sbjct: 301 AEDD 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2201

A DNA sequence (GBSx2320) was identified in *S.agalactiae* <SEQ ID 6795> which encodes the amino acid sequence <SEQ ID 6796>. This protein is predicted to be vacB protein (vacB). Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2966(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9399> which encodes amino acid sequence <SEQ ID 9400> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15366 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 338/780 (43%), Positives = 485/780 (61%), Gaps = 47/780 (6%)

Query: 4 AKAFPKLIKTIISNLESHRQL---RFDDNGSLSLQKKEAKKEITVRGLFRANKAGFGFL- 59
 A+ F +L+K + LE + R D G +K ++G A+ GF FL
 Sbjct: 36 AEEFKELVKALVALEEKGLIVRTRSDRYG-----IPEKMNLIKKGKISAHAKGFAFLL 87

Query: 60 SIDQDEDDMFIGKNDIAYAIDGDTVEAVVKKPADRLNGTAAEARVNVNIVERSLKTIVGKF 119
 D D+FI N++ A++GD V + + +G+ E V+ I+ER+++ +VG +
 Sbjct: 88 PEDTSLSDVFIPPNELNTAMNGDIVMVRINSQS---SGSRQEGTVIRILERAIQRRVVGTY 144

Query: 120 VLDDERPKYAGYIKSNQKINQKIYIRKEPV--VLDGTEIIVKVDIDKYPTRGHDYFVASV 177
 + G++ ++KI I+I K +G +++ V+ YP G V
 Sbjct: 145 T----ETRNFGFVIPDDKKITSDFIPKNGKNGAEGHKVV-VKLTSYP-EGRMNAEGEV 198

Query: 178 RDIVGHQGDVGIDVLEVSMDIVSEFPEDVIAEANAIPDAPTEKDLIGRVDLRQEVITFT 237
 I+GH+ D GID+L V+ + EFP D + +A++ PD EKDL R DLR +V T
 Sbjct: 199 ETILGHKNDPFGIDILSVIHKHGLPGFPPADAMEQASSTPDTIDEKDLKDRRDLRDQVIVT 258

Query: 238 IDGADAKDLDDAVHIKLLDNGHFELGVHIADVSYYVTEGSALNREALSRGTSVYVTDVRV 297
 IDGADAKDLDDAV + LD+G ++LGVHIADVS+YVTE S +++EAL RGTSVY+ DRV+
 Sbjct: 259 IDGADAKDLDDAVTVTKLDDGSYKLGVHIADVSHYVTENSPIDKEALERGTSVYLVDRVI 318

Query: 298 PMLPERLSNGICSLNPNLDRLTQSCIMEIDQNGRVVNHQITQSVINTTYRMTYTAVNDII 357
 PM+P RLSNGICSLNP +DRLT SC M I+ G+V H+I QSVI TT RMTY+ VN I+
 Sbjct: 319 PMIPHRLSNGICSLNPKVDRLTLSCMTINSQGVTEHEIFQSVIKTTERMTYSDVNKIL 378

Query: 358 A-GDEEICSEYESIVSSVQHMVTLHHTLEAMRTRRGALNFDTSEAKIMVNDKGMFVDIVI 416
 DEE+ +YE +V + M L L R RGA++FD EAK++V+D+G D+VI
 Sbjct: 379 VDDDEELKQKYEPLVPMFKDMERLAQILRDKMRDGAVDFFPKAEKVLVDDEGAVKDVVI 438

Query: 417 RNRGIAERMIESFMLAANETVAEHYARLKLFFIYRIHEEPKAEKLQKFIDYASVFGVQIQ 476
 R R +AE++IE FML ANETVAEH+ + +PFIYRIHEEP AEKLQKF+++ + FG ++

-2481-

Sbjct: 439 RERSVAEKLIEEFMLVANETVAEHFHMNVFFIYRIHEEPNAEKLQKFLFVTTFGYVVK 498

Query: 477 GTATKITQSALQDFMKKVQGGQFSEVLSMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTS 536
GTA I ALQ + V+ +P . V+S ++LRSM+QA+Y + GH+GL+ E+YTHFTS

5 Sbjct: 499 GTAGNIHPRALQSILDAVRDRPEETVISTVMLRSMQAKYDPSLGHFGLSTEFYTHFTS 558

Query: 537 PIRRYPDLLVHRMIRDY-DDKAMDKA--DHFANLIPEIATQTSSLERRAIDAERIVEAMK 593
PIRRYPDL+VHR+IR Y + +D+A + +A +P+IA TSS+ERRA+DAER + +K

10 Sbjct: 559 PIRRYPDLLVHRLIRTYLINGKVDEATQEKWAERLPDIAEHTSSMERRAVIDAERETDDLK 618

Query: 594 KAEYMEEVVGEEFEGVVASVVKFGMFVELPNTIEGLIHVTTL-PEYYHFNERTLTLOGEK 652
KAEYM + +GEEF+G+++SV FGMFVELPNTIEGL+HV+ + +YY F+E+ + GE+

Sbjct: 619 KAEYMLDKIGEEFDGMISSTVNFGMFVELPNTIEGLVHVFSMTDDYYRDEQHFAMIGER 678

15 Query: 653 SGKVFVRVGGQIKVKLIRSDKETGDDIDFDYLPSEDFIVEKVSXSREGRENRSKREHQHR 712
+G VFR+G +I VK++ +K+ +IDF+ + +G P R + +

Sbjct: 679 TGNVFRIGDEITVKVVDVNKDERNIDFEIV-----GMKGTFRPRELD---- 721

20 Query: 713 ISDRDNKNKNTSKKKASRKPKRNSDSKSHHKKDDRTTGSTKKKTCKPFYKGVAKKGQKRK 772
S R K ++K+ S + S K + T KKK K+ F +K +K+K

Sbjct: 722 -SSRSRKRGPARKRVQSTNTFPVSPAPS-EEKGEWFTKPKKKKKRGRFQNAKQKRKKK 779

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6797> which encodes the amino acid sequence <SEQ ID 6798>. Analysis of this protein sequence reveals the following:

25 Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0811(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 579/773 (74%), Positives = 664/773 (84%), Gaps = 22/773 (2%)

35 Query: 1 MAGAKAFPKLTIKISNLESHRQLRFPDDNGSLSLQKKEAKKEITVRGLFRANKAGFGFLS 60
MAGAK FP LIKTIS +ES LRF D+GSL+L+K+ KKKE TV+G+FRANKAGFGFL
Sbjct: 27 MAGAKHFPSLIKTISKMESQSLLRPSDDGSLALRKEREKKKEPTVQGVFRANKAGFGFLH 86

40 Query: 61 IDQDEDDMFIGNDIAIYDIDGTVAVVKKPADRLNGTAAEARVNIVERSLKTIVGKVF 120
+D++EDDMFIG+ND+ YDIDGTV VVKKPADRL GTAAEA+VV IV+RSLKT VG F+
Sbjct: 87 VDENEEDMFIGNDVGVAIDGTVVVKPADRLKGTAAEAKVVAIVDRSLKTAVGTFFI 146

45 Query: 121 LDDERPKYAGYIKSNQKINQKIYIRKEPVVLDGTEIIKVDIDKYPTRGHDFVASVRDI 180
LDD++PKYAGYI+SKNQKI QKIYI+KEPVVL GTETIKVDIDKYP RGHDFVASVRDI
Sbjct: 147 LDDDKPKYAGYIRSNQKIQQKIYIKKEPVVLKGTETIIVDIDKYPTRGHDFVASVRDI 206

Query: 181 VGHQGDVGIDVLEVLSEMDIVSEFPEDVIAEANAIPDAPTEKDLIGRVDLRQEVTFITDG 240
VGHQGDVGIDVLEVLSEMDIVSEFP +V+AEANAI +APT KDLIGRVDLRQE T TIDG

50 Sbjct: 207 VGHQGDVGIDVLEVLSEMDIVSEFPAEVLAEANAISEAPTAKDLIGRVDLRQETITIDG 266

Query: 241 ADAKDLDLDAVHIKLLDNGHFELGVHIADVSYVTEGSALNREALSRGTSVYVTDREVPEML 300
ADAKDLDLDA+HIKLLDNG++ELGVHIADVSYVTEGSAL++EA++RGTSVYVTDREVPEML
Sbjct: 267 ADAKDLDLDAIHIKLLDNGNYELGVHIADVSYVTEGSALDKAIAARGTSVYVTDREVPEML 326

55 Query: 301 PERLSNGICSLNPNLDRLTQSCIMEIDQNGRVVNHQITQSVINTTYRMTYTAVNDIAGD 360
PERLSNGICSLNPN+DRLTQS +MEI+ G VVN+QI QSVI TTYRMTY+ VND+IAGD
Sbjct: 327 PERLSNGICSLNPNLDRLTQSALMEINSQGHVVNYQICQSVIKTTYRMTYSTVNDMIAGD 386

60 Query: 361 EEICSEYESIVSSVQHMTLHHTLEAMRTRRGALNFDTSEAKIMVNDKGMPPDIVIRNRG 420
EE E+ SI V MV LH LEAMR++RGALNFDT EAKI+VNDKGMPPD+V+R RG
Sbjct: 387 EEALQEFASIAADDVTLMVALHRIEAMRSKRGALENFDTQEAKIIVNDKGMPPDVVLRQRG 446

65 Query: 421 IAERMIESFMLAANETVAEHYARLKLPIFYRIHEEPKAEKLQKFIDYASVFGVQIQGTAT 480
IAERMIESFMLAANE VAEH+A+ KLPIFYRIHEEPKAEKLQ+FDYAS FG+ IQGTA

-2482-

Sbjct: 447 IAERMIESFMLAANECVAEHFAKALPFTYRIHEEPKAEKLQQFIDYASTFGIHIQGTAN 506

Query: 481 KITQSALQDFMKKVQGPGEVLMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR 540
KI+Q ALQ FM KV+GQPG+EVLMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR

5 Sbjct: 507 KISQEALQAFMAKVEGQFGAEVLNMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR 566

Query: 541 YPDLVHRMIRDYDDKAMDKADHFANLIPEIATQTSSLERRAIDAERIVEAMKKAEYMEE 600
YPDLVHRM+R+Y+ + +K DHFA +IPE+AT +S LERRAIDAER+VEAMKKAEYM E

10 Sbjct: 567 YPDLVHRMVRREYNQPSQEKRDHFAQIIPELATSSSQLERRAIDAERVVEAMKKAEYMAE 626

Query: 601 YVGEEFEGVVASVVKFGMFVELPNTIEGLIHVTLTPEYYHFNERTLTLOGEKSGKVFRVG 660
YVGEEF+G+V+SVVKFG FVELPNTIEGL+H+T+LPEYYHFNERTL+LQGEKSGKV+VG

Sbjct: 627 YVGEEFDGIVSSVVKFGFFVELPNTIEGLVHITSLPEYYHFNERTLSLQGEKSGKVFKVG 686

15 Query: 661 QQIKVKLIRSDKETGDIIDFDYLPSPDFIVEKVS KSSREGRPNRSSKREHQHRISDRDNKN 720
Q I+VKL+++DKETGDIIDF+YLPSPDF+VEK+ S + R +R K+

Sbjct: 687 QPIRVKLVKADKETGDIIDFEYLPSPDFVVEKIKMSDKASRRDR-----RKS 732

Query: 721 KNTSKKKASRKPKRNSDSKSHHHKDDRTTGSTKKTCKPFYKGVAKKGQKRKS 773
+SK ++PK + +K T G TKK +KKPFYK AKK +++S

20 Sbjct: 733 SKSSKGTKKKEPKVAKAK-----TKGKTKKSGKPFYKEQAKKSRKRS 777

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2202

A DNA sequence (GBSx2321) was identified in *S.agalactiae* <SEQ ID 6799> which encodes the amino acid sequence <SEQ ID 6800>. This protein is predicted to be VacB homolog (smpB). Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2988(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23745 GB:AF052209 VacB homolog [Streptococcus pneumoniae]
Identities = 121/155 (78%), Positives = 139/155 (89%)

40 Query: 1 MVKGQGNVVAQNKKAHHDYTTIVETIEAGIVLTGTEIKSVRAARITLKDGYAQIKNGEAWL 60
M KG+G VVAQNKA HDYTIV+T+EAG+VLTGTEIKSVRAARI LKDG+AQ+KNGE WL

Sbjct: 1 MAKGEGKVVAQNKKARHDYTTIVDTLEAGMVLGTGTEIKSVRAARINLKDGFQVKNGEVWL 60

45 Query: 61 INVHITPYDQGNINWQDPDRTRKLLLLKKREIEKISNELKGTGMTLVPLKVYLKDGFAKVL 120
NVHI PY++GNIWNQ+P+R RKLLL K++I+K+ E KGTGMTLVPLKVY+KDG+AK+L

Sbjct: 61 SNVHIAPYEEGNIWNQEPERRRKLHLHKKIQKLEQETKGTGMTLVPLKVYIKDGYAKLL 120

50 Query: 121 LGLAKGKHDYDKRESIKRREQNRDIAROLKNYNSR 155
LGLAKGKHDYDKRESIKRREQNRDIAR +K N R

Sbjct: 121 LGLAKGKHDYDKRESIKRREQNRDIARVMKAVNQR 155

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6801> which encodes the amino acid sequence <SEQ ID 6802>. Analysis of this protein sequence reveals the following:

55 Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2918(Affirmative) < succ>

-2483-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 124/155 (80%), Positives = 145/155 (93%)

Query: 1 MVKGGQNVVAQNKKAHHDYTIIVETIEAGIVLTGTEIKSVRAARITLKDGYAQIKNGEAWL 60
 M KG+G+++AQNKKA HDY IVET+EAGIVLTGTEIKSVRAARI LKDG+AQIKNGEAWL
 10 Sbjct: 1 MAKGEGHILAQNKKARHDYHIVETVEAGIVLTGTEIKSVRAARIQLKDGFQIKNGEAWL 60

Query: 61 INVHITPYDQGNINWQDPDRTRKLLKREIEKISNELKGTGMTLVPLKVYLKDGFAKVL 120
 +NVHI P++QGNINW DP+RTRKLLKREI ++NELKG+GMTLVPLKVYLKDGFAKVL
 10 Sbjct: 61 VNVHIAPFEQGNINWADPERTRKLLKREITHLANELKSGMTLVPLKVYLKDGFAKVL 120

15 Query: 121 LGLAKGKHHDYDKRESIKRREQNRDIARQLKNYNSR 155
 +GLAKGKH+YDKRE+IKRR+Q RDI +Q+K+YN+R
 Sbjct: 121 IGLAKGKHEYDKRETIKRRDQERDIKQMKHYNAR 155

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 20 vaccines or diagnostics.

Example 2203

A DNA sequence (GBSx2322) was identified in *S.galactiae* <SEQ ID 6803> which encodes the amino acid sequence <SEQ ID 6804>. Analysis of this protein sequence reveals the following:

Possible site: 14
 25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.6876 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 35 vaccines or diagnostics.

Example 2204

A DNA sequence (GBSx2323) was identified in *S.galactiae* <SEQ ID 6805> which encodes the amino acid sequence <SEQ ID 6806>. This protein is predicted to be d-serine/d-alanine/glycine transporter (cycA). Analysis of this protein sequence reveals the following:

40 Possible site: 55
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -9.02	Transmembrane	71 - 87 (62 - 90)
INTEGRAL	Likelihood = -8.92	Transmembrane	320 - 336 (316 - 344)
INTEGRAL	Likelihood = -8.33	Transmembrane	254 - 270 (251 - 275)
45 INTEGRAL	Likelihood = -6.00	Transmembrane	158 - 174 (154 - 175)
INTEGRAL	Likelihood = -2.76	Transmembrane	197 - 213 (196 - 213)
INTEGRAL	Likelihood = -2.50	Transmembrane	117 - 133 (116 - 136)
INTEGRAL	Likelihood = -1.38	Transmembrane	282 - 298 (279 - 298)
50 INTEGRAL	Likelihood = -0.32	Transmembrane	342 - 358 (342 - 360)

----- Final Results -----
 bacterial membrane --- Certainty=0.4609 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-2484-

A related GBS nucleic acid sequence <SEQ ID 9397> which encodes amino acid sequence <SEQ ID 9398> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB14651 GB:Z99117 amino acid permease [Bacillus subtilis]
Identities = 165/361 (45%), Positives = 227/361 (62%), Gaps = 17/361 (4%)

Query: 1 MGIFLT-LSYWISLIFIGMAEITAVGEYVQFWFPEWPSWIIQIVFLAILSSINLIQVAVKAF 59
M F+T +YW I + MA++TAVG Y Q+W P+ P W+ ++ L IL +NL VK F
10 Sbjct: 95 MAAFITGWTYWFPCWISLAMADLTAVGIYTOYWLDPVQWLPGLLALIILLIMNLATVVKLF 154

Query: 60 GETEFWFAMIKVIAILGLIATGIFMVLTNFDTGHGYHASISNITNHFWEFPGKLNFFMA 119
GE EFWFA+IKVIAIL LI TGI ++ F G AS++N+ +H FP G F ++
15 Sbjct: 155 GELEFWFALIKVIAILLALIVTGILLIAGKFSASAG-PASLNNLWSHGGMFPNGWHGFIIS 213

Query: 120 FQMVFFAYLAIEFVGVTTSETANPRKVLPAKIQEIPMRIILFYAGSLLAIMAIFPQQQLP 179
FQMV FA++ IE VG+T ET NP+KV+PKAI +IP+RI+LFY G+L IM I+PW L
20 Sbjct: 214 FQMVVFAFVGIELVGLTAGETENPQKVIKAINQIPVRILLFYVGALFVIMCIYPWNVLN 273

Query: 180 VNESPFTVTFKLKAGIKWAAALINFFVLTSAASALNSTLYSTGRHLFQLANE--SPNALTK 237
NESPFTV VF GI AA+LINFVLTSAASA NS L+ST R ++ LA + +P L K
25 Sbjct: 274 PNESPFTVQVFSAVGIVVAASLINFVLTSAASAANSALFSTSRMVYSLAKDHHAPGLLKK 333

Query: 238 ALKLDQLSRQSVPSRAIIAS--AVIVGASALISVLPGISDAFSLITASSSGVYISIVVLI 295
L+ +VPS A+ S A+++G S L ++P F+LIT+ S+ +I I+ +
30 Sbjct: 334 -----LTSSNVPSNALFFSSIAILLIGVS-LNYLMP--EQVFTLITSVSTICFIFIWGIT 384

Query: 296 MIAHWKYRKS--PDFMEDGYKMPAYKILSPITLLFFLVFVSLFLQDSTYIGAIGATIWI 354
+I H KYR+ + + +KMP Y ++ +TL F F+ V L L + T I +W +
Sbjct: 385 VICHLYKRYKTRQHEAKANKFKMPFYPLSNYLTAFILVILALANDTRIALFVTPVWVF 445

There is also homology to SEQ ID 4070:

Identities = 286/364 (78%), Positives = 322/364 (87%), Gaps = 1/364 (0%)

35 Query: 2 GIFLTLSYWISLIFIGMAEITAVGEYVQFWFPEWPSWIIQIVFLAILSSINLIQVAVKAFGE 61
G F LSYWISLIFIGMAEITAVG YVQFWFP WP+W+IQ+VFL +LSSINLIQV+ FGE
Sbjct: 101 GYFGLSYWISLIFIGMAEITAVGAYVQFWFSPWPAWLQVLVLLSSINLIQVAVRVFGE 160

40 Query: 62 TEFWFAMIKVIAILGLIATGIFMVLTNFDTGHGYHASISNITNHFWEFPGKLNFFMAFQ 121
TEFWFAMIK++AIL LIAT IFMVL T F+T H HAS+SNI +HF FP GKL FFMAFQ
Sbjct: 161 TEFWFAMIKILAILALIATAIFMVLTFET-HTGHASLSNIFDHFSMFNGKLFMAFQ 219

45 Query: 122 MVFFAYLAIEFVGVTTSETANPRKVLPAKIQEIPMRIILFYAGSLLAIMAIFPQQQLPVN 181
MVFFAY AIEFVG+TTSETANPRKVLPAKIQEIP RI++FY G+L++IMAI PW QLPV+
Sbjct: 220 MVFFAYQAIEFVGITTTSETANPRKVLPAKIQEIPTRIVIFYVGALVSIMAI VFWQLPVD 279

50 Query: 182 ESPFVTVFKLAGIKWAAALINFFVLTSAASALNSTLYSTGRHLFQLANESPNALTKALKL 241
ESPFV VFKL GIKWAAALINFFVLTSAASALNSTLYSTGRHL+Q+ANE+PNALT LK+
Sbjct: 280 ESPFVMVFKLIGIKWAAALINFFVLTSAASALNSTLYSTGRHLYQIANETPNALTNRKI 339

55 Query: 242 DQLSRQSVPSRAIIASAVIVGASALISVLPGISDAFSLITASSSGVYISIVLIMIAHWK 301
+ LSRQ VPSRAIIASAV+VG SALI++LPG++DAFSLITASSSGVYI+IY L MIAHWK
Sbjct: 340 NTLSRQGVPSRAIIASAVVVGISALINILPGVADAFSLITASSSGVYIATYALTMIAHWK 399

60 Query: 302 YRKSPDFMEDGYKMPAYKILSPITLLFFLVFVSLFLQDSTYIGAIGATIWIIGFGLYSH 361
YR+S DFM DGY MP YK+ +P+TL FF FVF+SLFLQ+STYIGAIGATIWI FG+YS+
Sbjct: 400 YRQSKDFMADGYLMPKYVTTPTLTAFFAFVVISLFLQESTYIGAIGATIWIIFGIYSN 459

Query: 362 FKHK 365
K K
Sbjct: 460 VKFK 463

-2485-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2205

A DNA sequence (GBSx2324) was identified in *S.agalactiae* <SEQ ID 6807> which encodes the amino acid sequence <SEQ ID 6808>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have an uncleavable N-term signal seq

10	INTEGRAL	Likelihood = -8.33	Transmembrane	194 - 210 (191 - 215)
	INTEGRAL	Likelihood = -5.47	Transmembrane	17 - 33 (14 - 38)
	INTEGRAL	Likelihood = -5.15	Transmembrane	125 - 141 (119 - 144)
	INTEGRAL	Likelihood = -3.88	Transmembrane	155 - 171 (153 - 176)
	INTEGRAL	Likelihood = -1.38	Transmembrane	96 - 112 (94 - 114)
	INTEGRAL	Likelihood = -0.43	Transmembrane	49 - 65 (49 - 65)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.4333(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]

Identities = 80/214 (37%), Positives = 122/214 (56%), Gaps = 3/214 (1%)

25 Query: 4 FFSNIRTEIPQMPLLIHSLILSVLPFLMWLTLVNRDKPLYKTIWSILLGLQLITIYTWWF 63
FF+ T+ P+ L + + ++L + R+K +Y+ + IL +QLI +Y W++
Sbjct: 7 FFTTQATKPPKFDLFWVVSFLTLALTFYTAHRYREKQVYQRFQILQTVQLILLYGWYW 66

30 Query: 64 WAKLPLSESLPLYHCRIGMFVLLARPGI--LKDYFALLGVVGGVLAMHPDFYPYQFLH 121
+PLSESLP YHCR+ MFVLL PG K YFALLG G + A ++P Y F H
Sbjct: 67 VNHMPLSESLPFYHCRMAMFVLL--PGQSKYKQYFALLGTFGTAAAFVYPVPDAYPPH 125

35 Query: 122 VTNIFFIGHFALFVLSLLHMTQSNLDKLNPKLIITLTLINMSLIFINLLTGGNYGFM 181
+T + F GH AL SL++L+ Q N L+ K I +T +N + +NL+TGG+YGF+
Sbjct: 126 ITILSFIFGHLALLGNSLVLLRQYNARLLDVKGIFLMTFALNALIFVNLVTGGDYGFL 185

Query: 182 MKTPILGITNPFLNLFIVTTLLSFLVLFVKQIFQ 215

K P++G N +V+ +L + K+I +

Sbjct: 186 TKPPLVGDHGLVANYLLVSVILVATISLTKKILE 219

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6809> which encodes the amino acid sequence <SEQ ID 6810>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

45	INTEGRAL	Likelihood = -11.25	Transmembrane	16 - 32 (11 - 39)
	INTEGRAL	Likelihood = -3.45	Transmembrane	154 - 170 (153 - 173)
	INTEGRAL	Likelihood = -3.08	Transmembrane	96 - 112 (94 - 112)
	INTEGRAL	Likelihood = -1.91	Transmembrane	191 - 207 (191 - 209)
	INTEGRAL	Likelihood = -1.12	Transmembrane	71 - 87 (71 - 87)

50 ----- Final Results -----

bacterial membrane --- Certainty=0.5501(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]

Identities = 90/231 (38%), Positives = 128/231 (54%), Gaps = 7/231 (3%)

Query: 3 FFAIDPIGLPHTSLIFYLSSLLIALLLVFLTFQAYRLKS-HRYFFLFLQLSQVIGLYTWY 61

-2486-

FF P L +Y+S L L L F T YR K ++ FF LQ Q+I LY WY
 Sbjct: 7 FFTTQATKPPKFDLFWYVS-LFTLLALTFYTAHRYREKKVYQRFQILQTVQLILLVGYWY 65
 Query: 62 VLRGFPLDEALPLYHCRIMLAIFFLPDRNKFQLFMVLGIGGTFLALL--SPDLYPFRL 119
 + PL E+LP YHCR+AM + LP ++K+KQ F +LG GT A + PD YPF
 Sbjct: 66 WVNHMPLSESLPFYHCRMAMFVLLLPQSQSKYKQYFALLGTFGTLAAFPVPDAYPPF- 124
 Query: 120 WHVANVSFYFGHYALLVNGLIYLLRFYDASQLRLSVVRYLATVNFLLLLVSLATKGNYG 179
 H+ +SF FGH ALL N L+YLLR Y+A L + + +N L+ +V+L T G+YG
 Sbjct: 125 -HITILSFIFGHLALLGNLSVYLLRQYNARLLDVKGIFLMTFALNALIFVNVNLTGGDYG 183
 Query: 180 FVMDIPVIHTRHLLNFVIVTSGLTFMVKITEYFYLFKFGQAQQLALAFSKE 230
 F+ P++ L+ N+++V+ L + +T+ L+F AQ+ KE
 Sbjct: 184 FLTKPPLVGDHGLVANVYLLVSIIVLVAITSLTKKI-LEFFLAQEAEMIVKE 233

An alignment of the GAS and GBS proteins is shown below.

Identities = 70/216 (32%), Positives = 117/216 (53%), Gaps = 1/216 (0%)
 Query: 2 IEFFSNIRTEIPQMPLLIHSLILSVLPFLMWLTLVNRDKPLYKTIWSILLGLQITITYTW 61
 ++FF+ +P L+ + L + L++LT ++ + L Q+I +YTW
 Sbjct: 1 MDFFAIDPIGLPHTSLIFYLSSLLIALLLVFLTFQAYRLKSHRYFFLFLQLSQVIGLYTW 60
 Query: 62 FFWAKLPLSESLPLYHCRIGMFVVL-LARPGILKDYFALLGVVGGVLAMIHPDFYPYQFL 120
 + PL E+LPLYHCRIM + L K F +LG+ G LA++ PD YP++
 Sbjct: 61 YVLRGFPLDEALPLYHCRIMLAIFFLPDRNKFQLFMVLGIGGTFLALLSPDLYPFRLW 120
 Query: 121 HVTNIFFFIGHFALFVLSLLHMTQSNLDKLNPKLIQLTLLINMSLIFINLLTGGNYGF 180
 HV N+ F+ GH+AL V L++L+ + +L +++ +N L+ ++L T GNYGF
 Sbjct: 121 HVANVSFYFGHYALLVNGLIYLLRFYDASQLRLSVVRYLATVNFLLLLVSLATKGNYG 180
 Query: 181 MMKTPILGITNPFLNLFIVTTLSSFLVLFVKQIFQK 216
 +M P++ + LN IVT+ L+F+V + + K
 Sbjct: 181 VMDIPVIHTRHLLNFVIVTSGLTFMVKITEYFYLK 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2206

A DNA sequence (GBSx2325) was identified in *S.galactiae* <SEQ ID 6811> which encodes the amino acid sequence <SEQ ID 6812>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3297(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2207

A DNA sequence (GBSx2326) was identified in *S.galactiae* <SEQ ID 6813> which encodes the amino acid sequence <SEQ ID 6814>. This protein is predicted to be oxalate:formate antiporter (oxIT-2). Analysis of this protein sequence reveals the following:

-2487-

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

5	INTEGRAL	Likelihood = -7.80	Transmembrane	380 - 396 (376 - 399)
	INTEGRAL	Likelihood = -7.43	Transmembrane	291 - 307 (284 - 310)
	INTEGRAL	Likelihood = -5.63	Transmembrane	169 - 185 (163 - 186)
	INTEGRAL	Likelihood = -4.99	Transmembrane	226 - 242 (223 - 245)
	INTEGRAL	Likelihood = -4.19	Transmembrane	46 - 62 (39 - 63)
	INTEGRAL	Likelihood = -4.09	Transmembrane	311 - 327 (308 - 329)
10	INTEGRAL	Likelihood = -1.49	Transmembrane	261 - 277 (260 - 278)
	INTEGRAL	Likelihood = -1.06	Transmembrane	133 - 149 (133 - 150)
	INTEGRAL	Likelihood = -0.85	Transmembrane	98 - 114 (98 - 114)
	INTEGRAL	Likelihood = -0.06	Transmembrane	77 - 93 (77 - 93)

----- Final Results -----

15	bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20	>GP:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus lactis]	
	Identities = 220/398 (55%), Positives = 306/398 (76%), Gaps = 3/398 (0%)	
25	Query: 5	NRYVVAVSGVVLHMLGSTYAWSVFRNPIISETGWDISSVSFAFSLAIFCLGMSAAFMGH 64
		NRYVVA +GV+ HLM+GS YAWSVF NPI + GW SSV+ AFS+AI+ LGMSAAFMG
	Sbjct: 4	NRYVVAFAAGVMFHLMIGSVYAWSVFTNPIAKQNGWAESSVALAFSAIYFLGMSAAFMGK 63
30	Query: 65	LVERFGPRIMGMISSAILYGAGNVLTGLAIETQQLWLLYVAYGILGGIGLGSYITPVSTI 124
		+VE+ GPR+ G I++ LYG G ++TG AI +WLLY++YG++GG+GLG+GY+TPVSTI
	Sbjct: 64	VVEKIGPRLTGTIASFLYGTGTIMTGWAIHQNSIWLLYLSYGVIGGLGLGAGYVTPVSTI 123
35	Query: 125	IKWFPDRRGLATGFAIMGFGFASLVTSPLAQSLMTRIGVGKTFYILGLVYFFVMMIASQF 184
		IKWFPD+RGLATG AIMGFGFA+++T P+AQ LM +G+ +TFY+LG YF +M++A+QF
	Sbjct: 124	IKWFPDKRGLATGLAIMGFGFAAMLTGPVAQQLMASVGLBQTFYLLGTFFYFVIMLLAAQF 183
40	Query: 185	IKQPPQEKITILTHDGKKNAMNSQIITG--LKANAAIKSKTFYIIWLTFLFINISCGGLI 242
		I + P ++ T + +++ G L AN A+K+K+F +W+ FINI+CG+GL+
	Sbjct: 184	IVR-PNLALSSTTENSISQKKGTRITRGPILTANQALKTKSFTFLWIMFFINITCGIGLV 242
45	Query: 243	SAASPMQADLAGYSAESAALLVGVLFNGFGRLWASLSDYIGRPLTFIILFIVNFIMT 302
		SAASPMQ + G S ++AA++VG++G+PNGFGRL+WA+LSDYIGRP TF +FI++ +M
	Sbjct: 243	SAASPMQASMTGMSVQTAAIMVGIIGLFNGFGRLIWATLSDYIGRPATFSAIFILDIVML 302
50	Query: 303	SSLFLSFNAIVFAIAMSILMTCYGAGFSLLPAYLSDIFGTELATLHGYSLTAWAIAGLF 362
		S++ + ++F IA+ +LM+CYGAGFS++PAYL D+FGTEL +HGY LTAWA AG+
	Sbjct: 303	SAILIFKLPLLFVIALCLLMSCYGAGFSVIPAYLGDVFGTELGAHVGYVLTAWAAAGVV 362
	Query: 363	GPLLLSKTYSWGSYQLTLMVFGFLFGLLLSLYLRLK 400
		GPLLLS T+ ++Y LTL F + L LL+S ++++
	Sbjct: 363	GPLLLSLTHQLFHNVTLTAAFILIDLALLISFWIQR 400

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6815> which encodes the amino acid sequence <SEQ ID 6816>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

55	INTEGRAL	Likelihood = -12.95	Transmembrane	289 - 305 (282 - 321)
	INTEGRAL	Likelihood = -11.83	Transmembrane	376 - 392 (372 - 397)
	INTEGRAL	Likelihood = -8.55	Transmembrane	163 - 179 (160 - 189)
	INTEGRAL	Likelihood = -7.75	Transmembrane	227 - 243 (221 - 247)
60	INTEGRAL	Likelihood = -5.89	Transmembrane	44 - 60 (41 - 67)
	INTEGRAL	Likelihood = -1.38	Transmembrane	310 - 326 (309 - 327)
	INTEGRAL	Likelihood = -0.90	Transmembrane	353 - 369 (353 - 369)
	INTEGRAL	Likelihood = -0.37	Transmembrane	138 - 154 (138 - 154)
	INTEGRAL	Likelihood = -0.06	Transmembrane	98 - 114 (98 - 114)
65	INTEGRAL	Likelihood = -0.00	Transmembrane	259 - 275 (259 - 275)

-2488-

----- Final Results -----

5 bacterial membrane --- Certainty=0.6180 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus
 lactis]
 10 Identities = 222/399 (55%), Positives = 305/399 (75%), Gaps = 3/399 (0%)

Query: 3 KTKRYIIATAGILLHMLGSTYAWSVYRNPIIQTGWDQAPVAFSAIFCLGLSAAFM 62
 KT RY++A AG++ HLM+GS YAWSV+ NPI ++ GW ++ VA AFS+AI+ LG+SAAFM
 15 Sbjct: 2 KTNRYVVAFAGVMFHLMIGSVYAWSVFTNPIAKQNGWAESSVALAFSAIYFLGMSAAFM 61

Query: 63 GNLVEQYGPRLTGTVSAILYASGNMLTGLAIDRKEIWLILYIGYGVIGGLGLGAGYITPIS 122
 G +VE+ GPRLTGT+++ LY +G ++TG AI + IWLLY+ YGVIGGLGLGAGY+TP+S
 20 Sbjct: 62 GKVVEKIGPRLTGTIASFLYGTGTIMTGWAIHQNSIWLLYLSYGVIGGLGLGAGYVTPVS 121

Query: 123 TIIKWFPDKRGMATGFAIMGFGFASLLTSPIAQWLIETEGLVATFYLLGLIYLIVMLFAS 182
 TIIKWFPDKRG+ATG AIMGFGFA++LT P+AQ L+ + GL TFYLLG Y ++ML A+
 25 Sbjct: 122 TIIKWFPDKRGLATGLAIMGFGFAAMLTGPVAAQQLMASVGLEQTFYLLGTFFVIMLLAA 181

Query: 183 QLIIKPTAAEIAILDKKRLQ--NNSYLI EG--MTAKEALKTKSFYCLWVILFINITCGLGL 239
 Q I++P A + + Q + L G +TA +ALKTKSF LW++ FINITCG+GL
 30 Sbjct: 182 QFIVRPNLALSSTTENSISQKKGTRLTRGPELTANQALKTKSFTFLWIMFFINITCGIGL 241

Query: 240 ISVVPMAQDLTGMSPEMSAIVVGAMGIFNGFGRLVWASLSDYIGRRVTVILLFLVSIIM 299
 +S +PMAQ +TGMS + +AI+VG +G+ENGFGRL+WA+LSDYIGR T +F++ I+M
 35 Sbjct: 242 VSAASPMASQMTGMSVQTAAMVGIIGLFGNGFGRLIWATLSDYIGRPATFSAIFILDIVM 301

Query: 300 TISLIFAHSSLIFMISIATLMTCYGAGFSLIPPYLSDLFGAKELATLHGYILTAWAIAAL 359
 +++ L+F+I++ LM+CYGAGFS+IP YL D+FG KEL +HGY+LTAWA A +
 40 Sbjct: 302 LSAILIFKLPLLEFVIALCLLMSCYGAGFSVIPAYLGDVFGTKELGAVHGYVLTAWAAAGV 361

Query: 360 TGPMLLSITVEWTHNYLLTLCVFIVLYILGLMVALRLKK 398
 GP+LLS+T + HNY LTL FI++ +L L+++ +++
 45 Sbjct: 362 VGPLLLSLTHQLFHNVTLTAAFILIDLALLISFWIQR 400

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 252/400 (63%), Positives = 329/400 (82%), Gaps = 2/400 (0%)

Query: 1 MKNLNRYVAVSGVVLHMLGSTYAWSVFRNPIISETGWDISSVSFAFSLAIFCLGMSAA 60
 M+ RY++A +G++LHMLGSTYAWSV+RNPI+ ETGWD + V+FAFSLAIFCLG+SA
 45 Sbjct: 1 MEKTKRYIIATAGILLHMLGSTYAWSVYRNPIIQTGWDQAPVAFSAIFCLGLSAA 60

Query: 61 FMGHIVERFGPRIMGISAILYAGAGNVLTGLAIETQQLWLLYVAYGILGGIGLGSYITP 120
 FMG+LVE++GPR+ G +SAILY +GN+LTGLAI+ +++WLLY+ YG++GG+GLG+GYITP
 50 Sbjct: 61 FMGNLVEQYGPRLTGTVSAILYASGNMLTGLAIDRKEIWLILYIGYGVIGGLGLGAGYITP 120

Query: 121 VSTIIKWFPDRRGLATGFAIMGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMMI 180
 +STIIKWFPD+RG+ATGFAIMGFGFASL+TSP+AQ L+ G+ TFY+LGL+Y VM+
 55 Sbjct: 121 ISTIIKWFPDKRGMATGFAIMGFGFASLLTSPIAQWLIETEGLVATFYLLGLIYLIVMLF 180

Query: 181 ASQFIKQPQEKITILTHDGKKNAMNSQIITGLKANAAIKSKTFYIIWLTFLFINISCGLG 240
 ASQ I +P +I IL D K+ NS +I G+ A A+K+K+FY +W+ LFINI+CGLG
 60 Sbjct: 181 ASQLIIKPTAAEIAIL--DKKRLQNNSYLI EGMTAKEALKTKSFYCLWVILFINITCGLG 238

Query: 241 LISAASPMADLAGYSAESAALLVGVLGIFNGFGRLWASLSDYIGRPLTFIILFIVNFI 300
 LIS +PMAQDL G S E +A++VG +GIFNGFGRL+WASLSDYIGR +T I+LF+V+ I
 65 Sbjct: 239 LISVVPMAQDLTGMSPEMSAIVVGAMGIFNGFGRLVWASLSDYIGRRVTVILLFLVSIIM 298

Query: 301 MTSSFLFSFNAIVFAIAMSILMTCYGAGFSLPAYLSDFGTKEATLHGYSLTAWAIAAG 360
 MT SL + ++++F I+++ LMTCYGAGFSL+P YLSD+FG KELATLHGY LTAWAIA
 70 Sbjct: 299 MTISLIFAHSSLIFMISIATLMTCYGAGFSLIPPYLSDLFGAKELATLHGYILTAWAIAA 358

Lipop: Possible site: -1 Crend: 5
McG: Discrim Score: 5.06
GvH: Signal Score (-7.5): 4.38
 Possible site: 27

ALOM program count: 10 value: -7.80 threshold: 0.0

```

INTEGRAL    Likelihood = -7.80    Transmembrane    380 - 396 ( 376 - 399)
INTEGRAL    Likelihood = -7.43    Transmembrane    291 - 307 ( 284 - 310)
INTEGRAL    Likelihood = -5.63    Transmembrane    169 - 185 ( 163 - 186)
INTEGRAL    Likelihood = -4.99    Transmembrane    226 - 242 ( 223 - 245)
INTEGRAL    Likelihood = -4.19    Transmembrane    46 - 62 ( 39 - 63)
INTEGRAL    Likelihood = -4.09    Transmembrane    311 - 327 ( 308 - 329)
INTEGRAL    Likelihood = -1.49    Transmembrane    261 - 277 ( 260 - 278)
INTEGRAL    Likelihood = -1.06    Transmembrane    133 - 149 ( 133 - 150)
INTEGRAL    Likelihood = -0.85    Transmembrane    98 - 114 ( 98 - 114)
INTEGRAL    Likelihood = -0.06    Transmembrane    77 - 93 ( 77 - 93)
PERIPHERAL  Likelihood = 0.42      352
modified ALOM score: 2.06

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

ORF02272(313 - 1500 of 1818)
 GP|7107009|gb|AAF36228.1|AF168363_4|AF168363(4 - 400 of 421) oxalate:formate antiporter
 {Lactococcus lactis}
 %Match = 38.5
 %Identity = 55.4 %Similarity = 79.1
 Matches = 220 Mismatches = 81 Conservative Sub.s = 94

216 246 276 306 336 366 396 426
GK*IC*AENW*YIQFFDNLFITNYIFKNKT*VRF*EDCLKNLRVYVVAVSGVLHLMLGSTYAWSVFRNPPIISETGWDIS
 ||||| :||::||:|| ||||| ||| : || |
MKTNRVYVAFAGVMFHLMGISVYAWSVFNTNPIAKQNGWAES
 10 20 30 40

SVSFAFSLAIFCLGMSAAFMGHLVERFGPRIMGMISAILYAGNVLTGLAIETQQLWLLLYVAYGILGGIGLGSYGITPV
SVALAFSLAIYFLGMSAAFMGKVVKEIGPRLTGTIASFLYGTGTIMTGWAIHONSIWLLYLSYGVIGGLGLGAGYVTPVS

696 726 756 786 816 846 876 906
TIIKWFPPDRGLATGFAIMGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMMIASQFIKQPPOEKITILTHDGKK
|||||::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
TIIKWFPDKRGLATGLAIMGFGFAAMLTGPVAQCQLMASVGLGEQTFFYLGLTGFYFVIMLLAAQFIVRP-NLALSSTENSIS
 140 150 160 170 180 190 200

936 960 990 1020 1050 1080 1110 1140
NAMNSQIITG--LKANAATKSKTFYIIWLTLFINISCGLLISAASPMADLAGYSABSAALLVGVGLGFNGFGRLWAS
:: | | | :|:|:| :|: | |||:|:| |:| |||| | : | | :|:|:|:|:|:| |||:|:|
QKKGTRLTRGPFLTANQALKTKSFTEFLWMFFINITCGILGLVSAASPMAGSMVTGAATAIMVGII GLFNGFGRLIWAT
210 220 230 240 250 260 270 280

1170 1200 1230 1260 1290 1320 1350 1380

-2490-

```

      LSDYIGRPLTFTIILFIVNFIMTSSFLSFNAIVFAIAMSILMTCYGAGFSLLPAYLSDIFGTKELATLHGYSLTAWAIAAG
      ||||| | :||: : |::: :: ||: |:|:|||||)::||| |:|||| :|| |||||:|
      LSDYIGRPATFSAIFILDIVMLSAILIFKLPLLFFVIALCLLMSCYGAGFSVIPAYLGDFVTGKELGAHVHGYVLTAWAAAG
      290          300          310          320          330          340          350          360

5
      1410          1440          1470          1500          1530          1560          1590          1620
      LFGPLLLLSKTYSWGNSYQLTLMVFGFLFLFGLLLSLYLRLTKTVV*YISNLKFPGFTKEFFL*KIVLSYSK*FDILSI*
      : ||||| |:  ::| ||} } :: |: }:|):::::
      VVGPELLSLTHQLPHNYTLTLAAFILIDLALLISFWIQRFIKASKLIKKQIIKNYFKAH

10
      370          380          390          400          410          420

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2208

15 A DNA sequence (GBSx2327) was identified in *S.agalactiae* <SEQ ID 6817> which encodes the amino acid sequence <SEQ ID 6818>. This protein is predicted to be D-Ala-D-Ala adding enzyme (murF). Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1311(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9739> which encodes amino acid sequence <SEQ ID 9740> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC95436 GB:AF068901 D-Ala-D-Ala adding enzyme [Streptococcus pneumoniae]
Identities = 313/453 (69%), Positives = 375/453 (82%)

Query: 32 MKLSLHVAVKVGAKNQVSEFEDVPLGNIEFDSRNISEGDLFLPLKGARDGHEFIEMAFD 91
MKL++HE+A+VVGAKN+S FED L EFDSR I GDLF+PLKGARDGH+FIE AF+
Sbjct: 1 MKLTIHEIAQVVGAKNDISIFEDTQLEKAEFDSRLIGTGDLFVPLKGARDGHDFIETAFE 60

Query: 92 NGAIATISEKEIEGHPYLLVSDALKAFLVLAQYYIEKMNVVDIAVTGSNGKTTTKDMIAA 151
NGA T+SEKE+ HPY+LV D L AFQ LA YY+EK VDV AVTGSNGKTTTKDM+A
Subject: 61 NGAAVTLSEKEVSNHPYILVDDVLTAFLSLASYLEKTTVDVFAVTGSNGKTTTKDMLAH 120

Query: 152 ILSTTYKTYKTQGNYNNEIGLPYTVLHMPEDTEKIILEMGOHDLGDIHVLSEIAKPRIAV 211
+LST YKTYKTQGNYNNEIGLPYTVLHMPE TEK++LEMGODHLGDIH+LSE+A+P+ A+
Subject: 121 LLSTRYKTYKTQGNYNNEIGLPYTVLHMPEGTEKLVLEMGODHLGDIHLLSELARPKTAI 180

Query: 212 VTLIGEHAHLEFFGSREKIAEGMKQITDGMSSDGILIA PGDPIIDPYLPANQMTIRFGHDQ 271
VTL+GEAHL FF R +IA+GKMQL DGM+S +L+AP DPI++ YLP ++ +RFG
Sbjct: 181 VTLVGEAHLAFFKDRSEIAKKGKMOIADGMASGSLLLAPADPIVEDYLPPTDKKVVRFGGGA 240

Query: 272 ELQVTELKEEKSLTFKTNALEHQLRIPVPGKYATNAMVAAYVGKLLAVAEEDIVDALE 331
EL++T+L E K SLTFK N LE L +PV GKYNATNAM+A+YV V+EE I A +
Subject: 241 ELEITDLVERKDSLTFKANFLEOVLDLPTGKYATNAMTASYVALOEGVSEECIHOAFO 300

Query: 332 NLQLTRNRT EWKKSANGADILSDVYNANPTAMRLILETFSAIPNNDGGKKIALADMKE 391
+L+LTRNRT EWKK+ANGADILSDVYNANPTAM+LILETFSAIP N+GGKKIA+LADMKE
Sbjct: 301 DLRLTRNRT EWKKAANGADILSDVYNANPTAMKRLILETFSAIPANEGGKKIAVLADMKE 360

Query: 392 GEQSVDLHNNQMIMSIRPDSIDTLICYGQDIEGLAQLASQMFPIGKVVYFFKKNQEVDDQFDQ 451
G QSV LHNQMI+S PD +DT+I YG+DI LAQLASQMFPIG VY+FKK ++ DQF+
Subject: 361 GNQSVQLHNNQMILSLSPDVLDTVIFYGEDIAELAQLASQMFPIGHVYVYFKKTEDQDQFED 420

-2491-

Query: 452 LIAKVKDTLKEKDQILLKGSNSMNLISKIVDILE 484
 L+ +VK++L DQILLKGSNSMNL+ +V+ LE
 Sbjct: 421 LVKQVKESLSANDQILLKGSNSMNLAMLVESLE 453

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6819> which encodes the amino acid sequence <SEQ ID 6820>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 323/452 (71%), Positives = 387/452 (85%)

Query: 32 MKLSLHEVAKVVGAKNQVSEFEDVPLGNIEFDSRNISEGDLFLPLKGARDGHEFIEMAFD 91
 MKL+LHEVAK+V A+N VS+ +DVPL +IEFDSR I++GDLFLPLKG RDGHEFI++AF
 20 Sbjct: 1 MKLTLHEVAKIVDAQNNVSDLDVPLHHIEFDSRKITKGDLFLPLKGQRDGHFIDLAFO 60

Query: 92 NGAIATISEKEIEBHPYLLVSDALKAFQVLAQYYIEKMNVDVIAVTGSNGKTTTKDMIAA 151
 NGA+AT SEKE+ G P+LLV D LKAFQ LA YYI+KM VDVIAVTGSNGKT+TKDMI A
 25 Sbjct: 61 NGAVATFSEKELEGKPHLLVEDCLKAFQKLAHYIIDKMRVDVIAVTGSNGKTSTKDMIGA 120

Query: 152 ILSTTYKTYKTQGNYNNEIGLPYTVLHMPEDTEKIILEMGQDHLGDIHVLSEIAKPRIAV 211
 +LSTTYKTYKTQGNYNNEIGLPYTVLHMP+DTEKI+LEMGQDH+GDI +LSEIA+PRIAV
 30 Sbjct: 121 VLSTTYKTYKTQGNYNNEIGLPYTVLHMPDDTEKIVLEMGQDHMGDIRLLSEIARPRIAV 180

Query: 212 VTLLGEAHLEFFGSREKIAEGKMQITDGMSSDGILAPGDPIIDPYLPANQMTIRFGHDQ 271
 +TL+GEAHLE+FGSR+KIA+GKMQI DGM+SDGILAPGDPIIDPYLP NQM IRFG+ Q
 35 Sbjct: 181 LTLVGEAHLEYFGSRDKIAQGKMQIVDGMNSDGILAPGDPIIDPYLPENQMVIRFGNQ 240

Query: 272 ELQVTELKEEKHSLTFKTNALEHQLRIPVPGKYATNAMVAAYVGKLLAVABEDIVDALE 331
 E+ VT ++E+K SLTF TN L + +P+PGKYATNAMVAAYVGKLLAV +EDI+ AL+
 40 Sbjct: 241 EIDVTGIQEDKDSLTFTTNVLATPVSLPLPGKYATNAMVAAYVGKLLAVTDEDITIALQ 300

Query: 332 NLQLTRNRTEWKKANGADILSDVYNANPTAMRLILETFSAIPNNDGGKKIALADMKEL 391
 + LT NRTEWKK+ANGADILSDVYNANPTAMRLILETF+ I N GGKKIA+LADMKEL
 45 Sbjct: 301 TVTLTGNRTEWKKANGADILSDVYNANPTAMRLILETFANIAKNPGGKKI AVLADMKEL 360

Query: 392 GEQSVDLHNQMIMSIRPDSIDTLCYGDIEGLAQLASQMFPIGKVYFFKKNQEVDFDQ 451
 G+ SV LH+Q+I S+ +ID L+ YG I+ LA+LASQ++P +V++F K ++ DQF+
 50 Sbjct: 361 GKDSVILHSQLIDSLTSGNIDQLVFYGDHIKELARLASQVPAEQVHYFLKTEQEDQFEA 420

Query: 452 LIAKVKDTLKEKDQILLKGSNSMNLISKIVDIL 483
 + V++ L DQILLKGS+SM+L K+VD L
 Sbjct: 421 MAQYVQNILNPFQILLKGSMSLEKLVDR 452

- 50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2209

A DNA sequence (GBSx2328) was identified in *S.agalactiae* <SEQ ID 6821> which encodes the amino acid sequence <SEQ ID 6822>. Analysis of this protein sequence reveals the following:

55 Possible site: 17
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1381(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2492-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAC95435 GB:AF068901 D-Ala-D-Ala ligase [Streptococcus pneumoniae]
Identities = 243/346 (70%), Positives = 289/346 (83%)

Query: 3 KETLILLYGGRSAEREVSVLSAESVMRAINVDKFFVKTYFITQVGQFIKTQEFDEMPSSD 62
K+T+ILLYGGRSAEREVSVLSAESVMRA+NYD+F VKT+FI+Q G FIKTQEF P +
10 Sbjct: 2 KQTIILLYGGRSAEREVSVLSAESVMRAVNYDRFTVKTFFISQSGDFIKTQEFSHAPQOE 61

Query: 63 EKLMTNQTVDLDKMRPSDIYDDNAIVFPVLHGPMGEDGSIQGFLEVLMPYVGTNLS 122
++LMTN+T+D DK V PS IY++ A+VFPVLHGPMGEDGS+QGFLEVL+MPYVG NLS
15 Sbjct: 62 DRLMTNETIDWDKKVAPSAIYEEGAVVFPVLHGPMGEDGSVQGFLEVLKMPYVGCNLS 121

Query: 123 SVAMDKITTKQVLATVGVQVAYQTYFEGDDLEHAIKLSLETLSFPIFVKPANMGSSVGI 182
S+AMDKITTK+VL + G+ QV Y EGDD+ I E L++P+F KP+NMGSSVGI
15 Sbjct: 122 SLAMDKITTKRVLESAGIAQVPYVAIVEGDDVTAKIAVEBEKLAYPVFTKPSNMGSSVGI 181

Query: 183 SKATDESSLRSALDLALKYDSRLIEQGVTAAREIEVGILGNNDVKTTFPGEVVKDVFYD 242
SK+ ++ LR A+ LA +YDSR+L+EQGV AREIEVG+LGN DVK+T PGEVVKDV FYD
20 Sbjct: 182 SKSENQEELRQALKLAFRYDSRVLVEQGVNAREIEVGILLGNVDVKSTLPGEVVKDVAFYD 241

Query: 243 YDAKYIDNKITMDIPAKVDEATMEAMRQYASKAFKAIGACGLSRCDFFLTQDQIFLNEL 302
YDAKYIDNKITMDIPAK+ + + MRQ A AF+AIG GLSRCDF T G+IFLNEL
25 Sbjct: 242 YDAKYIDNKITMDIPAKISDDVVAVMRQNAETAFAIGGLGLSRCDFFYTDKGEIFLNEL 301

Query: 303 NTMPGFTQWSMYPLLWENMGLTYSDLIEKLVLAKEMFEKRESHLI 348
NTMPGFTQWSMYPLLW+NMG++Y +LIE+LV LAKE F+KRE+HLI
30 Sbjct: 302 NTMPGFTQWSMYPLLWDMGISYPELIERLVDLAKESFDKREHLI 347

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4559> which encodes the amino acid sequence <SEQ ID 4560>. Analysis of this protein sequence reveals the following:

Possible site: 23
35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1451(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 261/348 (75%), Positives = 306/348 (87%)

45 Query: 1 MSKETLILLYGGRSAEREVSVLSAESVMRAINVDKFFVKTYFITQVGQFIKTQEFDEMP 60
MSK+TL+LLYGGRSAEREVSVLSAESVMRA+NYDKF VKTYFITQ+GQFIKTQ+F E PS
Sbjct: 1 MSKQTLVLLYGGRSAEREVSVLSAESVMRAVNYDKFLVKTYFITQMGQFIKTQFSEKPS 60

Query: 61 SDEKLMTNQTVDLDKMRPSDIYDDNAIVFPVLHGPMGEDGSIQGFLEVLMPYVGTNLS 120
E+LMTN+T++L + ++PSDIY++ A+VFPVLHGPMGEDGSIQGFLEVLMPY+GTN++
50 Sbjct: 61 ESERLMTNETIELTQKIKPSDIYEEGAVVFPVLHGPMGEDGSIQGFLEVLMPYIGTNVM 120

Query: 121 SSSVAMDKITTKQVLATVGVQVAYQTYFEGDDLEHAIKLSLETLSFPIFVKPANMGSSV 180
SSS+AMDKITTK+VL ++G+PQVAY Y +G DLE + +L L+FPFVKPANMGSSV
55 Sbjct: 121 SSSIAMDKITTKRVLESIGIPQVAYTVYIDQDLEACLVELTARLTFFIFVKPANMGSSV 180

Query: 181 GISKATDESSLRSALDLALKYDSRLIEQGVTAAREIEVGILGNNDVKTTFPGEVVKDVF 240
GISKA + LR AI LAL YDSR+LIEQGV AREIEVG+LGN+ VK+T PGEV+KDVDF
60 Sbjct: 181 GISKAQTKVELRKAIQLALTYDSRVLIEQGVVAREIEVGILLGNDKVKSTLPGEVVKDVF 240

Query: 241 YDYDAKYIDNKITMDIPAKVDEATMEAMRQYASKAFKAIGACGLSRCDFFLTQDQIFLN 300
YDY AKY+DNKITM IPA VD++ + MR YA AFKA+G CGLSRCDFFLT+DQQ++LN
Sbjct: 241 YDYQAKYVDNKITMAIPADVDQSVITEMRSYAEVAFKALGGCGLSRCDFFLTQDQVYLN 300

Query: 301 ELNTPMGFTQWSMYPLLWENMGLTYSDLIEKLVLAKEMFEKRESHLI 348

-2493-

ELNTMPGFTQWSMYPLLWENMGL Y DLIE+LV LA+EMF++RESHLI
 Sbjct: 301 ELNTMPGFTQWSMYPLLWENMGLAYPDLLIEELVTLAQEMFDQRESHLI 348

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2210

A DNA sequence (GBSx2329) was identified in *S. agalactiae* <SEQ ID 6823> which encodes the amino acid sequence <SEQ ID 6824>. This protein is predicted to be recombination protein (recR). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2540(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44615 GB:U58210 RecM [Streptococcus thermophilus]
 Identities = 181/198 (91%), Positives = 189/198 (95%)

Query: 1 MLYPTPIAKLIDSFSLPGIGTKTATRLAFYTTIGMSDEDVNEFAKNLLAAKRELYCSVC 60
 MLYPTPIAKLIDSFSLPGIG KTATRLAFYTI MSDEDVN+FAKNLLAAKRELYCSVC
 Sbjct: 1 MLYPTPIAKLIDSFSLPGIGAKTATRLAFYTTISMSDEDVNDFAKNLLAAKRELYCSVC 60

Query: 61 GNLTDDDPCLICTDKTRDQSVILVVEDSKDVSAMEKIQEYNGLYHVLHGLISPMNGISPD 120
 G LTDDDPCLICTD+TRD++ ILVVEDSKDVSAMEKIQEY GLYHVL GLISPMNG+ PD
 Sbjct: 61 GRLTDDDPCLICTDETRDRTKILVVEDSKDVSAMEKIQEYRGLYHVLQGLISPMNGVGPD 120

Query: 121 DINLKSLITRLMDGQVTEVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180
 DINLKSLITRLMD +V EVI+ATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI
 Sbjct: 121 DINLKSLITRLMDSEVDEVIATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180

Query: 181 EYADEVTLLRAIENRTEL 198
 EYADEVTLLRAIENRTEL
 Sbjct: 181 EYADEVTLLRAIENRTEL 198

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6825> which encodes the amino acid sequence <SEQ ID 6826>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2652(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 180/198 (90%), Positives = 192/198 (96%)

Query: 1 MLYPTPIAKLIDSFSLPGIGTKTATRLAFYTTIGMSDEDVNEFAKNLLAAKRELYCSVC 60
 +LYPTPIAKLIDS+SKLPGIG KTATRLAFYTTIGMS+EDVN+FAKNLLAAKRELYCS+C
 Sbjct: 1 VLYPTPIAKLIDSYSKLPGIGIKTATRLAFYTTIGMSNEDVNDFAKNLLAAKRELYCSIC 60

Query: 61 GNLTDDDPCLICTDKTRDQSVILVVEDSKDVSAMEKIQEYNGLYHVLHGLISPMNGISPD 120
 GNLTDDDPCLICTD+RDQ+ ILVVED+KDVSAAMEKIQEY+G YHVLHGLISPMNG+ PD
 Sbjct: 61 GNLTDDDPCHICTDTSRDQTTILVVEDAKDVSAMEKIQEYHGYHVLHGLISPMNGVGPD 120

-2494-

Query: 121 DINLKSLITRLMDGQVTEVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180
 DINLKSLITRLMDG+V+EVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI
 Sbjct: 121 DINLKSLITRLMDGKVSEVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180

5 Query: 181 EYADEVTLLRAIENRTEL 198
 EYADEVTLLRAIENRTEL
 Sbjct: 181 EYADEVTLLRAIENRTEL 198

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 2211

A DNA sequence (GBSx2330) was identified in *S.agalactiae* <SEQ ID 6827> which encodes the amino
 acid sequence <SEQ ID 6828>. Analysis of this protein sequence reveals the following:

Possible site: 23
 15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3144(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 25 vaccines or diagnostics.

Example 2212

A DNA sequence (GBSx2331) was identified in *S.agalactiae* <SEQ ID 6829> which encodes the amino
 acid sequence <SEQ ID 6830>. This protein is predicted to be penicillin-binding protein 2b. Analysis of this
 protein sequence reveals the following:

Possible site: 52
 30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -13.69 Transmembrane 23 - 39 (17 - 46)

----- Final Results -----
 35 bacterial membrane --- Certainty=0.6477(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAC44614 GB:U58210 penicillin-binding protein 2b [Streptococcus thermophilus]
 Identities = 341/683 (49%), Positives = 477/683 (68%), Gaps = 12/683 (1%)

Query: 4 RKKRYRLTVKKQNASIPRRNLNLFIIIVLLFTVLILRLBQMGIQSQSFYMKKLTALTSTY 63
 ++K R ++ +I RR+ LLF ++ +LF +L RL MQ+ +SFY KKL + YT
 45 Sbjct: 18 KRKEKRANKPRKPVNISRRVYLLFGVVFVLFLLLFARLTVMQVYNKSFYTKKLEDNSKYT 77

Query: 64 VKESKARGQIFDAKGVLVENDERPTVAFSRGNNISSQSIKBLANKLSHYITLTEVASSD 123
 V+ + RGQIFDAKG+ L N + + F+R N +SS ++K +A +L+ +TLTE +D
 Sbjct: 78 VRIASERGQIFDAKGIALTTNQSKDVITFTFRSNLVSSDTMKVAERLATLVILTETKVTD 137

50 Query: 124 RAKRDYYLADKANYKKVVESLPDSKRYDKFGNHLAESTVYANAVAAPVSAINYSEDELK 183
 R KR++YLAD ANYK+VV LP+ K+ DKFGN LAE+T+Y NA+ AVP A++YSEDELK
 Sbjct: 138 RQKREFYLADSANYKRVNDLPNDKKTDKFGNKLAEATYNNAINAVPDEAVDYSEDELK 197

-2495-

Query: 184 VVALFNQMNATPTFGSVKLSTGELSDDQIKKLDADKKELLGISVTSNWHRRKKGTSLSDI 243
 +V +++ MNA F +V L T +L+ DQI + A +KEL GI V +W R +SLS +
 Sbjct: 198 IVYIYSHMNAVSNSFSTVILKTADLTPDQIAIVAQKQKELNGIRVAKDWERHTSDSSLSPL 257

5 Query: 244 LGTISTEKAGLPREEVKYLLKKGYSLNDRVGTSYLEKQYEDDLQGIQIRKVVVNKKGKV 303
 +G +S+ +AGLP+E+ K YLKKGY+LNDRVGTSYLEK+YE++LQG +R++ V+K+GKV
 Sbjct: 258 IGRVSSSEAGLPQEDAKDYLLKKGYALNDRVGTSYLEKEYEELQKHTVREITVDKEGKV 317

10 Query: 304 VSDNITQEGKSGRNLKLTIDLNQNKVESILKQYVGSSELSSGRASFSEGMVAIAIEPSTG 363
 SD I Q+G G NLKLTIDL++Q VE IL Q SE+S +A++SEGMVAV + TG
 Sbjct: 318 DSDKIIQKSGKGNLKLITIDLDQKGVEDILGQQLSSEISGNKATYSEGMVAVVMNADTG 377

15 Query: 364 KVLAMAGLKNDHC--NLVDDSLGTIAKNFTPGSVVKGATLSSGWENKVLRGNEVLYDQEI 421
 VLAMAG K++ G + D+LGTI FPGSVVKGATL++GW + + G++VL DQ I
 Sbjct: 378 AVLAMAGQKHEQGAQDFKADALGTITDVFTPGSVVKGATLTAGWRSGATYGDQVLTQPI 437

20 Query: 422 -----ANIRSWFT-RGLTPISAAQALEYSSNTYMVQVALRLMGQDYNTGDALTDGRGYQEA 475
 I SWFT +G I+A QALEYSSNTYMVQ+A++ +GQ Y G +L+ ++A
 Sbjct: 438 NIASSPPTISWFTDKGSRATTATQALEYSSNTYMVQIAIKRLGQYVPGMSLSTDNMEKA 497

25 Query: 476 MAKLRKTYGEYGLGVSTGLDLP-ESEGYVPGKYSLGTTLMESFGQYDAYTPMQLGQYIST 534
 M LR TY E+G+GVSTGLDLP ESEGY+P Y++ L E+FGQYD+YT +QL QY+++
 Sbjct: 498 MTLRLDTYAEFGMGVSTGLDLPGBSEGYIPKNYNVANVLTEAFQGYDSYTTIQLAQYVAS 557

30 Query: 535 IANNGNRLAPHVVSIDIYEGNDSNKFQALVRSITPKTLNKIAISDQELAIIEGFFYNVNS 594
 IAN G R+APH+V IY+ + L ++ + LNK+++ ++L IIQ+GF++VNS
 Sbjct: 558 IANGGKRVAPHIVGGIYDAGKNGSLGTLSTVDTRVLNKLSDSKQLGIIQGFHDVNS 617

35 Query: 595 GSGYATGTSMRGVNTTISGKTGTAEFTAKNVNGQTVSTYINLNAIAYDTNR---KIAVAVM 651
 GS ATG +M ++ ISGKTGTAEFT+A + +G +V+T NLNA+AY T + K+AV +M
 Sbjct: 618 GSSLATGKAMASSIIPISGKTGTAEFTYATDGSNGSVITVNLNAVAYATAKDGTKLAVGIM 677

Query: 652 YPHVTTDTTKSHQLVARDMIDQY 674
 YPH +K+HQ + +++ Y
 Sbjct: 678 YPHALDWKSKAHQNAVKAIMELY 700

A related GBS gene <SEQ ID 8997> and protein <SEQ ID 8998> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop Possible site: -1 Crend: 8
 McG: Discrim Score: -12.38
 GvH: Signal Score (-7.5): -5.9
 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -12.42 threshold: 0.0
 45 INTEGRAL Likelihood = -12.42 Transmembrane 23 - 39 (18 - 46)
 PERIPHERAL Likelihood = 4.56 355
 modified ALOM score: 2.98

50 *** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.5967(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50.5/71.3% over 683aa

Streptococcus
 thermophilus
 60 GP|1685112| penicillin-binding protein 2b Insert characterized

ORF02276(307 - 2322 of 2643)
 GP|1685112|gb|AAC44614.1||U58210(17 - 700 of 704) penicillin-binding protein 2b
 {Streptococcus thermophilus}
 65 %Match = 38.5

-2496-

%Identity = 50.4 %Similarity = 71.2

Matches = 342 Mismatches = 189 Conservative Sub.s = 141

```

5      108      138      168      198      228      258      288      318
NHGR*NS*LPTTCFRI**KIKPCFRILLR*II*SLYKKFRPSWLEFFIIYNILSVCKKPF*YNSSQSFYSKELMLNRKK
      :      ::      :      ::|
      MTSFWEKNSQKWKWKWRQKRK
      10      20

10     348      378      408      438      468      498      528      558
RYRLTVKKONASIPRRNLNLLFFIIXLLFTVLILRLQEQMIGQQSFYMKKLTALTSYTVKESKARGQIFDAKGVVLVFNDE
|      ::      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
EKRANKPRKPVNISRRVYLLFGVVFVLLFLLLFARLTVMQVYNKSFYTKKLEDNSKYTVRIASERGQIFDAKGIALTTNQS
      30      40      50      60      70      80      90      100

15     588      618      648      678      708      738      768      798
RPTVAFSRGNNTSSQSIKELANKLSHYITLTVEVSSDRAKRDYVLADKANYKICVVESLPDSKRYDKFGNHLAESTVYANA
:      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
KDVITFTRSNLVSSDTMKSVAERLATLVTLTETKVTDRQKREFVLADSANYKRVVNDLPNDKKTDFGNKLAETIYNNA
      110     120     130     140     150     160     170     180

20     828      858      888      918      948      978      1008     1038
VAAVPVSAINYSEDELKVVALFNQMNATPTFGSVKLSTGELSDDIQIKLDADKKELGISVTSNWHRRKKGTSLSLILGT
:      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
INAVFDEAVDYSEDELKIVYIYSHMNAVSNFSTVILKTADLTPDQIAIVAAKQKELNGIRVAKDWERHTSDSSLSPLIGR
      190     200     210     220     230     240     250     260

25     1068     1098     1128     1158     1188     1218     1248     1278
ISTEKAGLRPREVKYKYLKGYSLNDRVGTSYLEKQYEDDLQGIQIRKVVVNKKGVVSDNITQEGKSGRNLLKLTIDLNY
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
VSSSEAGLPQEDAKDYLLKGYALNDRVGTSYLEKEYEELQGHKTVREITVDKEGKVDSDKLIQKSGSKNNLLKLTIDLDF
      270     280     290     300     310     320     330     340

30     1308     1338     1368     1398     1428     1452     1482     1512
QNKVESILKQYYGSELSSGRASFSEGMVAVAIEPSTGKVLAMAGLKNDHG--NLVDDSLGTIAKNFTPGSVVKGATLSSG
| | | | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QKGVEDILGQQLSSEISGNKATYSEGMVAVVMNADTGAVLAMAGQKHEQGAQDFKADALGTITDVFTPGSVVKGATLTAG
      350     360     370     380     390     400     410     420

35     1542     1566     1587     1614     1644     1674     1704     1734
WENKVLRLGNEVLYDQ--EIAN--IRSWFT-RGLTPISAAQALEYSSNTYVMQVALRLMGQDYNITGDALTDGRYQEAMAK
| : : :| :| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
WRSgaiYGDQVLTDPINIASSPPITSWFTDKGSRAITATQALEYSSNTYVMQIAIKRLGQQYVPGMSLSTDNMEKAMTT
      430     440     450     460     470     480     490     500

40     1764      1821      1851      1881      1911      1941      1971
LRKTYGREYGLGVSTGLDLP-ESEGYVPGKYSLGTTLMESFGQYDAYTPMQLGQYISTIANNGNRLAPHVVSDIYEGNDSN
| | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
LRDITYAEFGMGVSTGLDLPGESEGYIPKNYNVANVLTEAFQYDSYTTTQLAQYVASIANGGKRVAPHIVGGIYDAGKNG
      510     520     530     540     550     560     570     580

45     2001      2031      2061      2091      2121      2151      2181      2211
KFAQLVRSITPKTLNKIAISDQELAIQEGFYNNVNSGSGYATGTSMRGNVTTISGKTGTAEYFAKNVNGQTVSTYNLNA
:      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
SLGTLSSSTVDTRVLNKLSDLSKQLGIIQQGFHDVNSGSSLATGKAMASSIIPISGKTGTAEYATDGSNGSVTTVNLNA
      590     600     610     620     630     640     650     660

50     2262      2292      2322      2352      2382      2412      2442
IAYDTNR---KIAYAVMYPHVTTDTTKSHQLVARDMIDQYISQFTGQ*ERTFECFTQHQLN*LTAFAQNYRV*VLKQOHV
:| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
VAYATAKDGTKLAVGIMYPHALDWKSKAHQNAVKALMELYQNTH
      670     680     690     700

```

SEQ ID 8998 (GBS292) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 9; MW 103kDa).

GBS292-GST was purified as shown in Figure 211, lane 7.

-2497-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2213

A DNA sequence (GBSx2332) was identified in *S.agalactiae* <SEQ ID 6831> which encodes the amino acid sequence <SEQ ID 6832>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2644(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB51328 GB:AJ131985 phosphoglyceromutase [Streptococcus pneumoniae]
 Identities = 219/230 (95%), Positives = 226/230 (98%)

Query: 1 MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIQAAGIEFDLAFTSVLKR 60
 MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLI+ AGI+FD A+TSVLKR

20 Sbjct: 1 MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIKEAGIKFDQAYTSVLKR 60

Query: 61 AIKTTNLALEAADQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120
 AIKTTNLALEA+DQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL

25 Sbjct: 61 AIKTTNLALEASDQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120

Query: 121 PPDMAKDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG 180
 PP+M +DDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG

30 Sbjct: 121 PPNMDRDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG 180

Query: 181 AHGNSIRALVKHIKQLSDDEIMDVEIPNFPPLVFEFDEKLNLVSEYYLGK 230
 AHGNSIRALVKHIK LSDDEIMDVEIPNFPPLVFEFDEKLN+VSEYYLGK

30 Sbjct: 181 AHGNSIRALVKHIKGLSDDEIMDVEIPNFPPLVFEFDEKLN+VSEYYLGK 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6833> which encodes the amino acid sequence <SEQ ID 6834>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2646(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 206/229 (89%), Positives = 214/229 (92%)

Query: 1 MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIQAAGIEFDLAFTSVLKR 60
 MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLI+ AGIEFDLAFTSVL R

50 Sbjct: 1 MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIKEAGIEFDLAFTSVLTR 60

Query: 61 AIKTTNLALEAADQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120
 AIKTTNLALE A QLWVP E KSWRLNERHYG LTGKNKAEAAEQF DEQVHIWRRSYDVL

55 Sbjct: 61 AIKTTNLALENAGQLWVPTEKSWRLNERHYGALTGKNKAEAAEQFCDEQVHIWRRSYDVL 120

Query: 121 PPDMAKDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG 180
 PP MAKDE+SAH DRRYA LD ++IPDAENLKVTLER+P+WE+KIAPAL DGKNVFG

55 Sbjct: 121 PPAMAKDDEYSAHKDRRYADLDPALIPDAENLKVTLERAMPYWEEKIAPALLDGKNVFG 180

Query: 181 AHGNSIRALVKHIKQLSDDEIMDVEIPNFPPLVFEFDEKLNLVSEYYLG 229

-2498-

AHGNSIRALVKHIK LSDDEIMDVEIPNFPPLVFE DEKLN+V EYYLG
 Sbjct: 181 AHGNSIRALVKHIKGLSDDEIMDVEIPNFPPLVFELDEKLNIVKEYYLG 229

SEQ ID 6832 (GBS110) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 5 extract is shown in Figure 38 (lane 8; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 10; MW 53.9kDa).

The GBS110-GST fusion product was purified (Figure 204, lane 5) and used to immunise mice. The
 resulting antiserum was used for Western blot (Figure 252A), FACS (Figure 252B), and in the *in vivo*
 passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS
 10 bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 2214

A DNA sequence (GBSx2333) was identified in *Sagalactiae* <SEQ ID 6835> which encodes the amino
 15 acid sequence <SEQ ID 6836>. This protein is predicted to be triosephosphate isomerase (tpiA). Analysis of
 this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 36 - 52 (36 - 52)
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 25

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC43268 GB:U07640 triosephosphate isomerase [Lactococcus
 lactis]
 Identities = 164/252 (65%), Positives = 202/252 (80%)
 30
 Query: 1 MSRKPFIAGNWKNMKNPEEAKAFIEAVASKLPSSSELVEAGIAAPALTLSTVLEAAKGSEL 60
 MSRKPFIAGNWKNMKN EA+AF+EAV + LPSS+ VE+ I APAL L+ + +GSEL
 Sbjct: 1 MSRKPFIAGNWKNMKNLTSEAQAFVEAVKNNLPSSDNVESVIGAPALFLAPMAYLRQGSSEL 60
 35
 Query: 61 KIAAQNSYFENSAGFTGENSPKVLAEEMGTDYVVIGHSERRDYFHETDQDINKKAKAIFAN 120
 K+AA+NSYFEN+GAFTGENSP + ++G +Y++IGHSERR+YFHETD+DINKKAKAIFA
 Sbjct: 61 KLAIAENSYFENAGFTGENSPAAIVDLGIEYIIIGHSERRDYFHETDEDINKKAKAIFAA 120
 40
 Query: 121 GLTPIIICGESLETYEAGKAVEFVGAQVSAALAGLSEEQVSSLVIAYEPIWAIGTGKSAT 180
 G TPI+CCGE+LET+EAGK E+V Q+ A LAGL+ EQVS+LVIAYEPIWAIGTGK+AT
 Sbjct: 121 GATPILCCGETLETFEAGKTAEWVSGQIEAGLAGLTAEQVSNLVIAYEPIWAIGTGKTAT 180
 45
 Query: 181 QDDAQNMCKAVRDVVAADFGQAVADKVRVQYGGSVKPNVAEYMACPDVDGALVGGASLE 240
 + A C VR V +G+ V++ VR+QYGGSVKPE + MA ++DGALVGGASLE
 Sbjct: 181 NEIADETCGVVRSTVEKLYGKEVSEAVRIQYGGSVKPEIEGLMAKENIDGALVGGASLE 240
 50
 Query: 241 AESFLALLDFVK 252
 A+SFLALL+ K
 Sbjct: 241 ADSFLALLEMYK 252

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6837> which encodes the amino acid
 sequence <SEQ ID 6838>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

-2499-

INTEGRAL Likelihood = -1.81 Transmembrane 36 - 52 (36 - 52)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 220/251 (87%), Positives = 237/251 (93%)

10 Query: 1 MSRKPF IAGNWK MNKNP EEA KAF IEAVASKLP SSELVEAGIAAPALT LSTVLEAAKGSEL 60
MSRKP IAGNWK MNKNP +EAKAF +EAVASKLP S++LV+ +AAPA+ L T +EAAK S L
Sbjct: 1 MSRKPI IAGNWK MNKNP QEAKAF EAVASKLPSTDLVDVAAPAVDLVTIEAAKDSVL 60

15 Query: 61 KIAAQNSYFENS GAFTGENSPKVLAEMGTDYVVIGHSERRDYFHETDQDINKKAKAIFAN 120
K+AAQN YFEN+GAFTGE SPKVLAEMG DYVVIGHSERRDYFHETD+DINKKAKAIFAN
Sbjct: 61 KVAAQNCYFENTGAFTGETSPKVLAEMGADYVVIGHSERRDYFHETDEDINKKAKAIFAN 120

20 Query: 121 GLTPIICCGESLETYEAGKAVEFVGAQVSAALAGLSEEQVSSLVIAYEPIWAIGTGKSAT 180
GLTPI+CCGESLETYEAGKAVEFVGAQVSAALAGLS EQV+SLV+AYEPIWAIGTGKSAT
Sbjct: 121 GLTPIVCCGESLETYEAGKAVEFVGAQVSAALAGLSAEQVASLVAYEPIWAIGTGKSAT 180

25 Query: 181 QDDAQNMCKAVRDVVAADFGQAVADKVRVQYGGSVKPENVAEYMACPDVDGALVGGASLE 240
QDDAQNMCKAVRDVVAADFGQ VADKVRVQYGGSVKPENV +YMACPDVDGALVGGASLE
Sbjct: 181 QDDAQNMCKAVRDVVAADFGQEVADKVRVQYGGSVKPENVKDYMACPDVDGALVGGASLE 240

Query: 241 AESFLALLDFV 251
A+SFLALLDF+
Sbjct: 241 ADSFLALLDFL 251

30

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2215

35 A DNA sequence (GBSx2334) was identified in *S.galactiae* <SEQ ID 6839> which encodes the amino acid sequence <SEQ ID 6840>. Analysis of this protein sequence reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3050(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAB41198 GB:U75481 elongation factor-Tu [Streptococcus mutans]
Identities = 44/45 (97%), Positives = 45/45 (99%)

Query: 1 MVMPGDNVTIEVELIHPIAVEQGTTFSIREGGRTVSGGIVSEIEA 45
MVMPGDNVTI+VELIHPIAVEQGTTFSIREGGRTVSGGIVSEIEA
50 Sbjct: 117 MVMPGDNVTIDVELIHPIAVEQGTTFSIREGGRTVSGGIVSEIEA 161

There is also homology to SEQ ID 1022:

Identities = 44/45 (97%), Positives = 44/45 (97%)

55 Query: 1 MVMPGDNVTIEVELIHPIAVEQGTTFSIREGGRTVSGGIVSEIEA 45
MVMPGDNVTI VELIHPIAVEQGTTFSIREGGRTVSGGIVSEIEA
Sbjct: 371 MVMPGDNVTINVELIHPIAVEQGTTFSIREGGRTVSGGIVSEIEA 415

-2500-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2216

A DNA sequence (GBSx2335) was identified in *S.agalactiae* <SEQ ID 6841> which encodes the amino acid sequence <SEQ ID 6842>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -2.66    Transmembrane    81 - 97 ( 80 - 97)
    INTEGRAL    Likelihood = -2.60    Transmembrane    18 - 34 ( 17 - 34)
10
    ----- Final Results -----
                bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2217

A DNA sequence (GBSx2336) was identified in *S.agalactiae* <SEQ ID 6843> which encodes the amino acid sequence <SEQ ID 6844>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence
25
    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.0596(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 2218

A DNA sequence (GBSx2337) was identified in *S.agalactiae* <SEQ ID 6845> which encodes the amino acid sequence <SEQ ID 6846>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence
40
    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.3559(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2501-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2219

A DNA sequence (GBSx2338) was identified in *S.galactiae* <SEQ ID 6847> which encodes the amino acid sequence <SEQ ID 6848>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have an uncleavable N-term signal seq

```

----- Final Results -----
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:AAF96286 GB:AE004374 hypothetical protein [Vibrio cholerae]
      Identities = 56/167 (33%), Positives = 89/167 (52%), Gaps = 12/167 (7%)

      Query: 18 LAIIKSLPLNDCWLCAGTLRNFWNKLS-GINETLTSDIDVFFDKNI---SYEETVVLE 73
      L + L L C++ AG +RN VW+ L + T +DIDV+FFD + YE++ LE
20      Sbjct: 41 LECVYQLQLPQCYIAAGFVRNLVWDSLHHNVKLTPLNDIDVIFFDADCLDSYDEKS--LE 98

      Query: 74 QQLKDNYPQYDWELKNEFYMNTHSPNTPKYTSSKDAISKFPKCTAVGARLDDRNQLELY 133
      +L + EQ +W+++KN+ M+ + + P Y S+ DA+S +PEK TAV R + ++ E
25      Sbjct: 99 LKLSEQMPQLNWQVKNQAKMHLQNGDNP-YQSTLDAMSYWPEKETAVAVRKVEHRYECI 157

      Query: 134 LPYGEIEILNFIVSPTPYFEEDLLRYNVYLKRVDDKKWNNIWPRLTI 180
      +G E + ++ P Y ++ RV K W +WP L I
30      Sbjct: 158 SAFGFESLFGQGFITHNP-----KRAYGIFENRVKSGWLAMWPNLRI 199

```

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2220

A DNA sequence (GBSx2339) was identified in *S.galactiae* <SEQ ID 6849> which encodes the amino acid sequence <SEQ ID 6850>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45      >GP:CAB13060 GB:Z99110 yjdF [Bacillus subtilis]
      Identities = 47/138 (34%), Positives = 93/138 (67%), Gaps = 2/138 (1%)

      Query: 1 MKMTVYFDGNFWLGLIEYDDGDYKVFYFFGKEPKDDVFNFINHKLNDLIKKEFVK 60
      MK+T+Y+DG FW+G++E D+G + FR+ FGKEP+D +V F++++L +++ + E +
50      Sbjct: 24 MKLTIYYDGGQFWGVVEVVDNGKLRAFRHLFGKEPRDSEVLEFVHNQLLNMAQAE--QE 81

      Query: 61 DISLKRTNEHKKSPKRMQREINREKRKPVVSTKAQLAMKTIHMSIKNERQLSQKCKKNEL 120
      + L+ + K +PKR+QR++++E + V++KAQ A+K + K +++ K ++ +
55      Sbjct: 82 GVRQLGRRQKKINPKRLQROVSKELKNAGVTSKAQEAIKLELEARKQKKQIMKEQREHV 141

```

-2502-

Query: 121 RKHRYQLKQEKRYQKKKG 138
 ++ RY LK++K +K +G
 Sbjct: 142 KEQRYMLKKQKAKKKHRG 159

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2221

10 A DNA sequence (GBSx2340) was identified in *S.agalactiae* <SEQ ID 6851> which encodes the amino acid sequence <SEQ ID 6852>. This protein is predicted to be ComX1. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3143(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 9469> which encodes amino acid sequence <SEQ ID 9470> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAD50429 GB:AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 61/152 (40%), Positives = 95/152 (62%)
 Query: 5 EELFDKVKPIVMKLRNRYFVQLWEYDDWIQGRIVLFRLLLEHPYLLDNESKLFYFKTK 64
 +EL+++V+ V K R Y++ LWE DW QEG + L L+ L+D+ +L YFKTK
 Sbjct: 3 KELYEEVQGTIVYKCRNEYYLHLWELSDWDQEGMLCLHELISREGLVDDIPRLRKYFKTK 62
 30 Query: 65 FSNYLDNVLRHQDCQKRQFNKMPYEEISEVSHYVSKGLVLDDYIAYRDTLTKVEETLSD 124
 F N + D +R Q+ QKR+++K PYEE+ E+SH + GL LDDY + +TL S
 Sbjct: 63 FRNRILDYIRKQESQKRRYDKEPYEEVGEISHRISSEGGWLDDYYLPHETLRDYNKQSK 122
 35 Query: 125 IDKEKFEKLISGERFAGKKQFIRDIQFFNAF 156
 +E+ E+++S ERF G+++ +RD++ F F
 Sbjct: 123 EKQEELERVLNSNERFRGRQVLRDLRIVFKEF 154

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6853> which encodes the amino acid sequence <SEQ ID 6854>. Analysis of this protein sequence reveals the following:

40 Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-10.35 Transmembrane 9 - 25 (7 - 28)
 45 ----- Final Results -----
 bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 A related sequence was also identified in GAS <SEQ ID 9163> which encodes the amino acid sequence <SEQ ID 9164>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-10.35 Transmembrane 2 - 18 (1 - 18)

-2503-

----- Final Results -----

5 bacterial membrane --- Certainty=0.160(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAD50429 GB:AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 60/149 (40%), Positives = 98/149 (65%)

Query: 41 FEKVKPIILKLRHYIQLWDRDDLQEGHIIILLQLLERYPELIEEEERLYRYFKTKFSS 100
 +E+V+ + K + YY+ LW+ DW QEG + L +L+ R L+++ RL +YFKTKF +
Sbjct: 6 YEEVQGTIVYKCRNEYYLHLWELSDWDQEGMLCLHELISRBEGLVDDIPRLRKYFKTKFRN 65

15 Query: 101 YLKDLLRRQESQKRQPHKLAYEEIGEVAHAIPSRGLWLDYVAYQEVASLENQLNSQER 160
 + D +R+QESQKR++ K YEE+GE++H I GLWLDDY + E + N+ + +++
Sbjct: 66 RILDYIRKQESQKRRYDKPEYEEVGEISHRISEGLWLDYVAYQEVASLENQLNSQER 125

20 Query: 161 MQFQALIRGERFKGRALLRKISPYFKEF 189
 + + ++ ERF+GR+ +LR + FKEF
Sbjct: 126 EELERVLSNERFRGRQVRDLRIVFKEF 154

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 78/149 (52%), Positives = 116/149 (77%)

Query: 8 FDKVKPIVMKLRNRYFVQLWEYDDWIQEGRIVLFRLLLEHPYLLDNESKLFITYFKTKFSN 67
 F+KVKPI++KL+R+Y++QLW+ DDW+QEG I+L +LLE +P L++ E +L+ YFKTKFS+
Sbjct: 41 FEKVKPIILKLRHYIQLWDRDDLQEGHIIILLQLLERYPELIEEEERLYRYFKTKFSS 100

30 Query: 68 YLNDVLRHQDCQKRQFNKMPYEEISEVSHYVKSGLVLDYIAYRDTLTKEETLSIDK 127
 YL D+LR Q+ QKRQF+K+ YEEI EV+H + S+GL LDDY+AY++ + +E L+ ++
Sbjct: 101 YLKDLLRRQESQKRQPHKLAYEEIGEVAHAIPSRGLWLDYVAYQEVASLENQLNSQER 160

35 Query: 128 EKFEKLISGERFAGKKQFIRDIQPFNF 156
 +F+ LI GERF G++ +R I P+F F
Sbjct: 161 MQFQALIRGERFKGRALLRKISPYFKEF 189

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 **Example 2222**

A DNA sequence (GBSx2341) was identified in *S.agalactiae* <SEQ ID 6855> which encodes the amino acid sequence <SEQ ID 6856>. Analysis of this protein sequence reveals the following:

45 Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.23 Transmembrane 166 - 182 (166 - 182)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1893(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAA99510 GB:Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 57/226 (25%), Positives = 97/226 (42%), Gaps = 22/226 (9%)

Query: 4 VRLYIARHGKTMFNTIGRAQGSWDTPLTTFGELGIKELGLKASNISFKEAFSSDSGRT 63
 +RL+I RHG+T N QG DT + GE +LG L++ I F + SSD R
Sbjct: 17 IRLFIIRHGQTEHNVKKILOGHKDOTSINPTGEEQATKLGHYLSRGIHFDKVVSSDLKRC 76

-2504-

Query: 64 LQTMETIILREVQQENIPYTRDKRIREWCFGSLDGGYDGLFNGVLPRVSNMGMSHLTHEE 123
 QT ++L+ +QEN+P + +RE G ++G M E+

Sbjct: 77 RQTTALVLKHSKQENVPTSYSGLRERYMGVIEG-----MQITEAEK 118

5 Query: 124 IANLICQVDTAGWAEPPWAILSNRILSGFTAIAKKIEDIGGGNAIVVSHGMTIATFL-WL- 181
 A+ + + E R+ + + G N +VSHG I L WL

Sbjct: 119 YADKHGEGSFRNFGEKSDDFVARLTGCVEEVVAEASNEGVKNLALVSHGGAIRMIQLWLK 178

10 Query: 182 IDHSTPRSLGLDNGSVSVVDF--EDGTFSIQSIGDMSYREKGREIL 225
 ++ + + N SV++VD+ + F ++ +G+ + G ++

Sbjct: 179 YENHQAHKIIVFNTSVTIVDYVKDSKQFIVRRVGNTQHLGDGEFVV 224

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6857> which encodes the amino acid sequence <SEQ ID 6858>. Analysis of this protein sequence reveals the following:

15 Possible site: 57

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.69 Transmembrane 170 - 186 (170 - 186)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:CAA99510 GB:Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 64/231 (27%), Positives = 98/231 (41%), Gaps = 27/231 (11%)

30 Query: 5 RLYIARHGKTMFNTIGRAQGWSDTPLTKKGEEGIRELGGLKDATIPFKAAPSSDSGRIM 64
 RL+I RHG+T N QG DT + GEE +LG L+ I F SSD R

Sbjct: 18 RLFIRHGQTEHNVKKILQGHKDT SINPTGEEQATKLGHYLRSGIHFDKVVSSDLKRCR 77

Query: 65 QTIEIILRESENEFLPYTKDNRIREWCFGSLGTYDSEFLGVLPRTKAFENRDNLRDVP 124
 QT ++L+ S+ E +P + + +RE G +EG +E

35 Sbjct: 78 QTTALVLKHSKQENVPTSYSGLRERYMGVIEGMQITEA----- 116

Query: 125 YSELAESIVEVDTANWAEPPWEVLKRRIWEGFEAIALSIQNAAGGNALVSHGMTIGTFL- 183
 + A+ E N+ E + R+ E N G N +VSHG I L

40 Sbjct: 117 -EKYADKHGEGSFRNFGEKSDDFVARLTGCVEEVVAEASNEGVKNLALVSHGGAIRMIQL 175

Query: 184 WL--IDPDRDKQYIDNGSVTVVVF--DDGQFTIKTIGDMSYRYRGREIIEE 230
 WL + K + N SVT+V++ D QF ++ +G+ + G ++ +

Sbjct: 176 WLKYENHQAHKIIVFNTSVTIVDYVKDSKQFIVRRVGNTQHLGDGEFVVSD 226

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 150/231 (64%), Positives = 182/231 (77%), Gaps = 5/231 (2%)

Query: 1 MSKVRLYIARHGKTMFNTIGRAQGWSDTPLTTFGELGIKELGLGLKASNISFKEAFSSDS 60
 M+K RLYIARHGKTMFNTIGRAQGWSDTPLT GE GI+ELGLGLK + I FK AFSSDS

50 Sbjct: 1 MTKTRLYIARHGKTMFNTIGRAQGWSDTPLTKKGEEGIRELGGLKDATIPFKAAPSSDS 60

Query: 61 GRTLQTMETIILREVQQENIPYTRDKRIREWCFGSLDGGYDGLFNGVLPRV----SNGDM 116
 GRT+QT+ETIILRE + E +PYT+D RIREWCFGSL+G YD +LF GVLPR + ++

55 Sbjct: 61 GRTMQTIEIILRESENEFLPYTKDNRIREWCFGSLGTYDSEFLGVLPRTKAFENRDNL 120

Query: 117 SHLTHEETIANLICQVDTAGWAEPPWAILSNRILSGFTAIAKKIEDIGGGNAIVVSHGMTIA 176
 + + E+A I +VDTA WAEPP +L RI GF AIA I++ GGGNA+VSHGMTI

Sbjct: 121 RDVPYSELAESIVEVDTANWAEPPWEVLKRRIWEGFEAIALSIQNAAGGNALVSHGMTIG 180

60 Query: 177 TFLWLIDHSTPRSLGLDNGSVSVVDFEDGTFSIQSIGDMSYREKGREILEK 227
 TFLWLID + +DNGSV+V+V+DG F+I++IGDMSYR +GREI+E+

Sbjct: 181 TFLWLIDPDRDKQY-IDNGSVTVVVFDDGQFTIKTIGDMSYRYRGREIIEE 230

-2505-

A related GBS gene <SEQ ID 8999> and protein <SEQ ID 9000> were also identified. Analysis of this protein sequence reveals the following:

Cytoplasmic predicted but experimentally found on the surface of *Streptococci*

```

5      32.3/52.0% over 184aa

                                     Thermotoga maritima
      EGAD|165681| phosphoglycerate mutase Insert characterized
      GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase Insert characterized
      PIR|G72260|G72260 phosphoglycerate mutase - (strain MSB8) Insert characterized
10
      ORF01265(268 - 870 of 1248)
      EGAD|165681|TM1374(1 - 185 of 201) phosphoglycerate mutase {Thermotoga maritima}
      GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase {Thermotoga maritima}
      PIR|G72260|G72260 phosphoglycerate mutase - Thermotoga maritima (strain MSB8)
15      %Match = 6.3
      %Identity = 32.2 %Similarity = 52.0
      Matches = 57 Mismatches = 78 Conservative Sub.s = 35

      105      135      165      195      225      255      285      315
20      RGRNNSYEIFNPFMSMLLKRINRFYFCSR*LQNFFIGKVR*YIPVKAFVFCYNIKCL*GVSMKSVRLYIARHGKTMFNTII
                                     ::||: |||:|::|
                                     MKLYLIRHGETIWNK
                                     10

25      345      375      405      435      465      495      519      549
      GRAQGWSDTPLTTFGELGIKELGLKASNISFKEAFSSDSXRTIQTMEIILREVQQENI--PYTRDKRIREWCFCGSLDG
      |  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      GLWQGVTDVPLNERGREQARKLANSLK----RVDAIYSSPLKRSLETABEIIARRFEKEIIVEEDLRECEISLW-----
                                     30      40      50      60      70      80

30      579      609      639      669      699      729      759
      GYDGLDFNGVLPRVSNMGDMSHLTHEEIANLICQVDTAGW-----EPWAILSNRILSGFTAIAKKIEDIGGGNAI
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      -----NGLTVEE-AIREYPVEFKKWSSDPNFGMEGLSMRNVQNRVVKAIMKIVSQEKLNGSENVV
                                     90      100      110      120      130      140

35      789      816      840      870      900      930      960      990
      VVSHGMTIATFL-WLIDHST--PRSLGLDNGSVSVDFEDGTFISIQSIGDMSYREKGREILEKTLQ*KKIKLSDSV*LVF
      :  |  |  ::  |  :  ::  |  :  ||  |  :  ||  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
40      IVSHSLSLRAFIWILGLPLYLHRNFKLDNASLSVVEIESKPRLVLLNDTCHLKES
                                     160      170      180      190      200

```

SEQ ID 9000 (GBS44) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 6; MW 27kDa), in Figure 168 (lane 8-10; MW 42kDa – thioredoxin fusion) and in Figure 238 (lane 7; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 8; MW 52.4kDa).

Purified Thio-GBS44-His is shown in Figure 244, lanes 7 & 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 2223

A DNA sequence (GBSx2342) was identified in *S.agalactiae* <SEQ ID 6859> which encodes the amino acid sequence <SEQ ID 6860>. This protein is predicted to be d-alanyl-d-alanine carboxypeptidase. Analysis of this protein sequence reveals the following:

```

55      Possible site: 27
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----

```

-2506-

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 108/169 (63%), Positives = 139/169 (81%)

10 Query: 79 ELSPDVVPVENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFSYVTQEM 138
 E++PDV ++ + +D RI + +FL AA+ IDS EHLISGYRSVAYQE+L+N+Y+ QE
 Sbjct: 4 EMNPDVTDIDGVKVDSDRIAENTRKFLAAQAEIDSSEHLISGYRSVAYQEELYNNYIAQEK 63

15 Query: 139 TSNPNLTRGQAEKLVKTYSPAGASEHQTGLAMDSTVDSLNE SDPRVVSQ LKKIAPQYG 198
 +NP+L++ +A+K V+TYSQP G+SEHQTGLA+DMSTVDSL N+SD VV+++ IAP+YG
 Sbjct: 64 ANNPSLSQEEAQKQVQTYSPPGSSEHQTGLAIDMSTVDSL NQSDANVVAIAAPKYG 123

20 Query: 199 FVLRFPDGKTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKE 247
 FVLRFP+GK TG+ YEDWHYRYVG+SAKYM KH LTL EEY+ LKE
 Sbjct: 124 FVLRFPPEGKIDATGIDYEDWHYRYVGKSAKYMTKHDLTLEEYLKLLKE 172

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6861> which encodes the amino acid sequence <SEQ ID 6862>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 25 INTEGRAL Likelihood = -9.66 Transmembrane 10 - 26 (3 - 29)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 118/173 (68%), Positives = 139/173 (80%)

35 Query: 74 ITKEMSPELADINGISVDKRIEQATSDFLAAQAIDLQEHLISGYRSVDYQTELYQSYIK 133
 IT EM+P++ DI+G+ VD RI + T FLAAQ ID EHLISGYRSV YQ ELY +YI
 Sbjct: 1 ITAEMNPDVTDIDGVKVDSDRIAENTRKFLAAQAEIDSSEHLISGYRSVAYQEELYNNYIA 60

40 Query: 134 KEMANDPTLTQEAEEALVQTYSPPGASEHHTGLAIDMSTVDTLN ASDPSVAKAVQKIAP 193
 +E AN+P+L+QE A+ VQTYSPPG+SEH TGLAIDMSTVD+LN SD +V V IAP
 Sbjct: 61 QEKANNPSLSQEEAQKQVQTYSPPGSSEHQTGLAIDMSTVDSL NQSDANVVAIAIAP 120

45 Query: 194 DYGFVLRFPPEGKKTSTGVDYEDWHYRYVGKASARYMAQHNLTL EEYIAALKEK 246
 YGFVLRFPPEGK +TG+DYEDWHYRYVG SA+YM +H+LTL EEY+ LKEK
 Sbjct: 121 KYGFVLRFPPEGKIDATGIDYEDWHYRYVGKSAKYMTKHDLTLEEYLKLLKEK 173

An alignment of the GAS and GBS proteins is shown below.

Identities = 131/235 (55%), Positives = 172/235 (72%), Gaps = 3/235 (1%)
 50 Query: 15 LLAILCF--SLFALLKPNSQQSSSQKL RNE DIKKISSQKR NKLQ LPAVSSKDWNLILVN 72
 LL ++ F L+ +KP + +Q L ++I++ +K ++ LP VS +DW L+LVN
 Sbjct: 12 LLIVIVFLGGLYLF IKPEESVTPTQ-LNKKEIQKDIKKTDR LRALPKVSVEDWELVLVN 70
 55 Query: 73 RDHKHELSPDVVPVENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFS 132
 RDH +E+SP++ + I +DKRI + + FL AA+AID +EHLISGYRSV YQ +L+ S
 Sbjct: 71 RDHITKEMSPELADINGISVDKRIEQATSDFLAAQAIDLQEHLISGYRSVDYQTELYQS 130
 60 Query: 133 YVTQEMTSNPNLTRGQAEKLVKTYSPAGASEHQTGLAMDSTVDSLNE SDPRVVSQ LKK 192
 Y+ +EM ++P LT+ AE LV+TYSQP GASEH TGLA+DMSTVD+LN SDP V ++K
 Sbjct: 131 YIKKEMANDPTLTQEAEEALVQTYSPPGASEHHTGLAIDMSTVDTLN ASDPSVAKAVQK 190
 Query: 193 IAPQYGFVLRFPDGKTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKE 247

Sbjct: 191 IAP YGFVLRFP+GK TGV YEDWHYRYVG SA+YMA+H+LTLEEYI LKE 245

5

10

15

20

```

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

25

EGAD|41322| d-alanyl-d-alanine carboxypeptidase Insert characterized
GP|1209528|gb|AAB05624.1|U35369 D,D-carboxypeptidase Insert characterized

30

35

Matches = 63 Mismatches = 79 Conservative Sub.s = 40

40

45

50

55

60

SEQ ID 6860 (GBS18) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 3; MW 31kDa).

-2508-

The GBS18-His fusion product was purified (Figure 93A; see also Figure 189, lane 11) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 93B), FACS (Figure 93C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

5 Example 2224

A DNA sequence (GBSx2343) was identified in *S.galactiae* <SEQ ID 6863> which encodes the amino acid sequence <SEQ ID 6864>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```

Possible site: 34
10 >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood =-12.58    Transmembrane    10 - 26 ( 3 - 29)

----- Final Results -----
15         bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
         bacterial outside --- Certainty=0.0000(Not Clear) < succ>
         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6865> which encodes the amino acid sequence <SEQ ID 6866>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 33
    >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood =-11.83    Transmembrane    10 - 26 ( 4 - 33)

----- Final Results -----
25         bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
         bacterial outside --- Certainty=0.0000(Not Clear) < succ>
         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

30 >GP:AAD00279 GB:U78599 putative N-acetyl-muramidase [Streptococcus mutans]
    Identities = 66/150 (44%), Positives = 97/150 (64%), Gaps = 5/150 (3%)

Query: 18 LLLIVCPLLSSQRIASADKEVRVNYSQKQFITKMGKEVKPLAKYYGIRPSILIAQILLET 77
          LL+I+ P+L+S +A A+K++ YS K+F+ ++ + L+K YG+R SI+I Q L++
35 Sbjct: 3 LLVILLPILASGLADANKMPSPYSHKEFVKETAPTAKLSKIYGVRSIIIGQAALDS 62

Query: 78 HDGKTLTLLASKYHNLFSSKATPGQVAITLKSPPKQTN---QNV--RYAIYKDDASAIRDYLR 132
          H G TLLASKYHNLF S +A+PGQ A+ LKS + N Q V RY +Y+ ++ DY+
40 Sbjct: 63 HFGSTLLASKYHNLFSEASPGQGAIVRLKSHEYKNGRWQEVINRYLVYESWKESLYDYMA 122

Query: 133 MLRQGKEVDKRLYRNLTATEKGYKAPAKSLQ 162
          +L K DK LY + T GYK A++LQ
Sbjct: 123 ILHGNKINDKALYTTMTSSGYKTVARALQ 152

```

45 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 67/190 (35%), Positives = 102/190 (53%), Gaps = 1/190 (0%)

Query: 1 MRKRFSLLNFIVVTFIFFFIFLPLLNHKGKVDANSRQSVTYTKKEEFIQKIVPDAQDLGK 60
          MRKR F+ + F I+ PLL+ + A+ V Y++++FI K+ + + L K
50 Sbjct: 1 MRKRLKFPYFLTLLACFLLLVLCPLLSSQRIASADKEVRVNYSQKQFITKMGKEVKPLAK 60

Query: 61 SYGIRPSFIIAQALDSDFGEKILANKYHNLFGLLAEPGTPSITLNDSSSTGKKQEKQFTH 120
          YGIRPS +IAQ L++ G+ +LA+KYHNLF A PG +ITL S Q ++
Sbjct: 61 YYGIRPSILIAQILLETHDGKTLTLLASKYHNLFSSKATPGQVAITLK-SPKQTNQNVRYAI 119
55 Query: 121 YKSWKYSMDYLAHIKSGATGKKDSYTIMVSVKNPKTLVQKLQDSGFDNDKKYAKKMTET 180
          YK ++ DYL ++ G K Y + + K K + LQ DK YA+++ ++

```

-2509-

Sbjct: 120 YKDDASAIRDYLRMLRQGKEVDKRLYRNRLATEKGYKAPAKSLQKYLHYTDKTYARRLIQV 179

Query: 181 IDLYDLTRYD 190
I+ DLT YD

5 Sbjct: 180 IESNDLTNYD 189

SEQ ID 6864 (GBS246) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 7; MW 24.6kDa).

GBS246d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lanes 14 & 15; MW 21kDa) and in Figure 183 (lane 4; MW 21kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 12; MW 46kDa). Purified GBS246d-GST is shown in Figure 243, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 2225

A DNA sequence (GBSx2344) was identified in *S.agalactiae* <SEQ ID 6867> which encodes the amino acid sequence <SEQ ID 6868>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2541(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45610 GB:U78296 repressor of class I heat shock gene

expression HrcA [Streptococcus mutans]

Identities = 227/345 (65%), Positives = 287/345 (82%), Gaps = 1/345 (0%)

Query: 17 VITQRQNDILNLIVLELFTQTTHPEVGSKALQRTIDSSSATIRNDMAKLEKLGLEKAHTSS 76
+ITQRQ DIILNLIVLELFT+THEP+GSK LQ +I SS ATIRNDMA LEKLGLEKA T

Sbjct: 1 MITQRQNDILNLIVLELFTKTTHPEIGSKTLQNSIASSRATIRNDMAALEKLGLEKATTPP 60

Query: 77 GRM-PSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMQLKASHILSEMTGYT 135
+ P +YFVEHSL DS+DEQD+Y VIKAFDFEAF+L D+LQ+AS +L+ +TGYT

Sbjct: 61 AVVCPVKKAIRYFVEHSLNPDSDLDEQDVYQVIKAFDFEAFRLGDLQASDVLANLTGYT 120

Query: 136 SVILDVEPARQRLTGFDVVQLSNHDAVMTLDESKFVTVQFAIPRNFLTRDLIAFKAIV 195

++ILDVEP +QRLT FD+V+LSNHDAV+TLDE+ PVTVQFAIP+NFL DL+ I

Sbjct: 121 ALILDVEPKQRLTTFDIVKLSNHDAVLTLDSEFVTVQFAIPKFLDSLMTVAKIA 180

Query: 196 EERLLDGSMVDIHYKLRTTEIPQIVQKYFVTTDNVQLFDFYVFSELFLETVFVAGKVNSLT 255

ER L+ +V+DIHY+LRTE PQI+QKYF TDNVL LFD++F+ +F E VF++GK+ +L

Sbjct: 181 RERFLNQTVLDIHYRLRTEPPQIIQKYFPRTDNVLDLFDHIFNPIFQEEVFISGRIKTL 240

Query: 256 YSDLSTYQFLDNEQQVAISLRQSLKGEEMASVQVADSQEAALADVSVLTHKFLIPYRGFG 315

++ L TYQFL+N Q VA+ +RQSL E E+ VQVADS+E +LAD++V++ KFLIPYRGFG

Sbjct: 241 FAGLDTYQFLENLQSVALEIRQSLPEDELHRVQVADSKEKSLADLTVISQKFLIPYRGFG 300

Query: 316 LLSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDIYRYLNSNHYEVH 360

+L++IGP+D+DY+R++SL+N+I +VLA KLGD+YRYLNSNHYEVH

Sbjct: 301 ILTVIGPVDLDYQRTISLINVISRVLAVKLGDFYRYLNSNHYEVH 345

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6869> which encodes the amino acid sequence <SEQ ID 6870>. Analysis of this protein sequence reveals the following:

-2510-

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0695(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 341/344 (99%), Positives = 343/344 (99%)

Query: 17 VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLEKAHTSS 76
 VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLEKAHTSS
 15 Sbjct: 1 VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLEKAHTSS 60

Query: 77 GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILSEMTGYTS 136
 GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHIL+EMTGYTS
 20 Sbjct: 61 GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILAEEMTGYTS 120

Query: 137 VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNLTRDLIAFKAIVE 196
 VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNLTRDLIAFKAIVE
 25 Sbjct: 121 VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNLTRDLIAFKAIVE 180

Query: 197 ERLLDGSMVDIHYKLRTFIPQIVQKYFVTTDNVLQLEFDYVFSELFLETVFVAGKVNSLT 256
 ERLLD SV+DIHYKLRTFIPQIVQKYFVTTDNVLQLEFDYVFSELFLETVFVAGKVNSLT
 30 Sbjct: 181 ERLLDNSVIDIHYKLRTFIPQIVQKYFVTTDNVLQLEFDYVFSELFLETVFVAGKVNSLT 240

Query: 257 SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL 316
 SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL
 35 Sbjct: 241 SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL 300

Query: 317 LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYRYLNSNHYEVEH 360
 LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYRYLNSNHYEVEH
 40 Sbjct: 301 LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYRYLNSNHYEVEH 344

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2226

40 A DNA sequence (GBSx2345) was identified in *S.agalactiae* <SEQ ID 6871> which encodes the amino acid sequence <SEQ ID 6872>. This protein is predicted to be grpe protein (grpE). Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.5138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45611 GB:U78296 GrpE [Streptococcus mutans]

Identities = 130/180 (72%), Positives = 151/180 (83%), Gaps = 3/180 (1%)

55 Query: 14 VSEEEKDDLQEEVEATE--TEETVEEVIEEIEPKSELELANERADEFENKYLRAHAEM- 70
 +S++ KK++ +EEVEATE TEE+VEEV EE E EL+ A ERA++FENKYLRAHAEM
 Sbjct: 1 MSKKDKKEEYKEEVEATEPTTEESVEEVAEETSSENKELQEALERAEDFENKYLRAHAEMP 60

Query: 71 QNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGLEMTRDSLIQA 130
 + + QRYRSQDL KAILPSLDNLERALAVEGLTDDVKKGLEM ++SLIQA
 60 Sbjct: 61 KTFVALMKSDKVCQRYRSQDLRKAILPSLDNLERALAVEGLTDDVKKGLEMVQESLIQA 120

-2511-

Query: 131 LKEEGVEEVEVDSFDHNFHMAVQTL PADDEHPADSIAEVFQKGYKLHERLLRPAMVVVYN 190
 LKEEGVEEVE+++FD N HMAVQTL ADD+HPADSIA+V QKGY+LHERLLRPAMVVVYN
 Sbjct: 121 LKEEGVEEVELENFDANLHMAVQTL DADDDHPADSIAQVHQKGYQLHERLLRPAMVVVYN 180

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6873> which encodes the amino acid sequence <SEQ ID 6874>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 189/190 (99%), Positives = 189/190 (99%)

Query: 1 MAVFNKLFKRRHSVSEEEKDDLQEEVEATETETVEEVIEEIPKSELELANERADEF 60
 MAVFNKLFKRRHSVSEEEKDDLQEEVEATETETVEEVIEE PEKSELELANERADEF
 Sbjct: 1 MAVFNKLFKRRHSVSEEEKDDLQEEVEATETETVEEVIEETPEKSELELANERADEF 60
 Query: 61 NKYLRAHAEMONIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120
 NKYLRAHAEMONIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL
 Sbjct: 61 NKYLRAHAEMONIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120
 Query: 121 EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTL PADDEHPADSIAEVFQKGYKLHERL 180
 EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTL PADDEHPADSIAEVFQKGYKLHERL
 Sbjct: 121 EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTL PADDEHPADSIAEVFQKGYKLHERL 180
 Query: 181 LRPAMVVVYN
 LRPAMVVVYN
 Sbjct: 181 LRPAMVVVYN 190

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2227

A DNA sequence (GBSx2346) was identified in *S.agalactiae* <SEQ ID 6875> which encodes the amino acid sequence <SEQ ID 6876>. This protein is predicted to be heat shock protein 70 (dnaK). Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0996(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6877> which encodes the amino acid sequence <SEQ ID 6878>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0996(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2512-

An alignment of the GAS and GBS proteins is shown below.

Identities = 594/609 (97%), Positives = 603/609 (98%), Gaps = 1/609 (0%)

```

5  Query: 1  MSKIIGIDLGTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN 60
    Sbjct: 1  MSKIIGIDLGTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN 60

10 Query: 61  PDTVISIKSMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120
    Sbjct: 61  PETVISIKSMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120

15 Query: 121  NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE 180
    Sbjct: 121  NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE 180

20 Query: 181  LGDGVFDVLATAGDNKLGDDDFDQKIIDFLVEEFKKENGIDLSQDKMALQRLKDAAEKAK 240
    Sbjct: 181  LGDGVFDVLATAGDNKLGDDDFDQKIIDFLV EFKKENGIDLSQDKMALQRLKDAAEKAK 240

25 Query: 241  KDLGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS 300
    Sbjct: 241  KDLGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS 300

30 Query: 301  LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV 360
    Sbjct: 301  LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV 360

35 Query: 361  LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD 420
    Sbjct: 361  LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD 420

40 Query: 421  NKTILGRFQLTDIPAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVIQSNSGLTDE 480
    Sbjct: 421  NKTILGRFQLTDIPAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVI+SN GL++E 480

45 Query: 481  EIDKMMKDAEANAADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDERDAAQSALD 540
    Sbjct: 481  EIDKMMKDAEANAADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDERDAAQSALD 540

50 Query: 541  ELKKAQESGNLDDMKAKLEALNEKAQALAVKLYEQAAAAQQAQGAEGAQSADSSSKGDD 600
    Sbjct: 541  ELK KAQESGNLDDMKAKLEALNEKAQALAVK+YEQAAAAQQAQGAEGAQ+ DS++ DD 599

    Query: 601  VVDGEFTEK 609
    Sbjct: 600  VVDGEFTEK 608
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 2228

A DNA sequence (GBSx2347) was identified in *S.agalactiae* <SEQ ID 6879> which encodes the amino acid sequence <SEQ ID 6880>. This protein is predicted to be *Streptococcus pneumoniae* DnaJ protein homologue (dnaJ). Analysis of this protein sequence reveals the following:

```

55 Possible site: 18
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.4180(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
60    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

-2513-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6881> which encodes the amino acid sequence <SEQ ID 6882>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1322 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 330/377 (87%), Positives = 357/377 (94%), Gaps = 1/377 (0%)

Query: 1 MNNTIEFYDRLGVSKDASQDEIKKAYRRMSKKYHPDINKETGAEKYKEVQEAYETLSDTQ 60

MNNTIE+YDRLGVSKDASQD+IKKAYR+MSKKYHPDINKE GAE+KYK+VQEAYETLSQ+Q

Sbjct: 19 MNNTIEFYDRLGVSKDASQDDIKKAYRKMSKKYHPDINKEAGAEQKYKDVQEAYETLSDSQ 78

Query: 61 KRAAYDQYGAAGANGGFGGFDGGGFGGFDGGGFGGFEDIFSSFFGGGMRPNAPRQGDD 120

KRAAYDQYGAAGA GGFGG GGFGGFDGGGFGGFEDIFSSFFGGGG RPNAPRQGDD

Sbjct: 79 KRAAYDQYGAAGACGGFGG-GAGGFGGFDGGGFGGFEDIFSSFFGGGSRPNAPRQGDD 137

Query: 121 LQYRVNLSFEEAIFGAEEKVSYNRESSCHTCSGSGAKPGTSPVTCQKCHGSGVINVDQT 180

LQYRVNLSFEEA+FG EKEVSYNRE++C TC GSGAKPGT+PVTC+KCHGSGV+ +DTQT

Sbjct: 138 LQYRVNLSFEEAVFGVEKEVSYNREATCGTCLGSGAKPGTAPVTCRKCHGSGVMTIDTQT 197

Query: 181 PLGTMRRQVTCDCVCGSGQEIKEKPCPTCHGTGHEKKTHKVSVKIPAGVETGQQIRLTGQG 240

PLG MRRQVTCDC+ C GSG+EIKE C TCHGTGHEK+ HKVSVKIPAGVETGQQIRL GQG

Sbjct: 198 PLGMMRRQVTCDCICHGSGKEIKEPCQTCHGTGHEKQAHKVSVKIPAGVETGQQIRLTGQG 257

Query: 241 EAGFNGGFPYGDLEFVI+NVLPSSQFERNGSTIYYTLNISFVQAALGDTIDIPTVHGAVEMS 300

EAGFNGGFPYGDLEFVI+NVLPSS+QFERNGSTIYY L+ISF QAALGDT++IPTVHG VEM+

Sbjct: 258 EAGFNGGFPYGDLEFVILNVLPSSQFERNGSTIYYNLDISFTQAALGDTVEIPTVHGDEVEMA 317

Query: 301 IPAGTQTGKTFRRLRGKGAPKLRGGGQGDQHVTVNIVTPTKLNDQAQKEALHAFEAESGDKM 360

IPAGTQTGKTFRRL+GKGAPKLRGGGQGDQHVTVNIVTPTKLNDQAQ+EAL AFAEASG+KM

Sbjct: 318 IPAGTQTGKTFRRLRGKGAPKLRGGGQGDQHVTVNIVTPTKLNDQAQREALQAFEAESGERM 377

Query: 361 VHPKKKGFFDKVKDALD 377

+HPKKKGFFDKVKDAL+

Sbjct: 378 LHPKKKGFFDKVKDALE 394

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2229

A DNA sequence (GBSx2348) was identified in *S.agalactiae* <SEQ ID 6883> which encodes the amino acid sequence <SEQ ID 6884>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.22 Transmembrane 281 - 297 (281 - 297)

----- Final Results -----

bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD24445 GB:AF118389 unknown [Streptococcus suis]

Identities = 182/373 (48%), Positives = 257/373 (68%), Gaps = 5/373 (1%)

-2514-

Query: 4 KVEEIRSYLIASIQNGKLAPGDRLPISIRQLANQFSCNKDITVQVLMELRFDNYIYAKPRS 63
 K + I ++ I+ + G++LPSIRQL Q+ C+KDTVQ+ ++EL++ N IYA +S
 Sbjct: 3 KYQVIIQDILTGIEEHRFKRGEKLPISIRQLREQYHCSKDITVQKAMLELKYQNKIYAVEKS 62

5 Query: 64 GYVFDSSHQEEVEEGVSLPNSEIANIAYDDFRLCLNETLIGREDYLFNYYYRQEGLLDLS 123
 GYY+ + + + + ++ I Y+DFR+CL E+LIGRE+YLFNYY++QEGL +L
 Sbjct: 63 GYVILEDRDFQ-DHTCRAQSYRLSRITYEDFRICLKESLIGRENYLFNYYHQEGLAELI 121

10 Query: 124 KAVAKLMEETGVYVPLDDIVITAGTQQALFILTQVTFPNRKSRLVIEEPTYPRMIELIKT 183
 +V L+ + VY D +VITAG+QQAL+ILTQ+ K+ +LIE PTY RMIELI+
 Sbjct: 122 SSVQSLMDYHVTYTKKQDLVITAGSQQALYILTQMETLAGKTEILIENTPYSRMIELIRH 181

15 Query: 184 QNLPHYETISRGTGIDFQRLLEEIFQTQSIKFFYVIPRMHNPLGTSYNPVMKRLIEMAOK 243
 Q +PY+TI R GID + LE IFQT IKFFY IPR+HNPLG++Y+ ++++A++
 Sbjct: 182 QGIPYQTIERNLDGIDLEBLESIFQTGKIKFFYTIPRLHNPLGSTYDIATKTAIVKLAKQ 241

20 Query: 244 YDVYIVEDDYMSDFASQS--PLHYDTHGRVIYLSFSKAIFFPALRLAAICLPQALKSTF 301
 YDVYI+EDDY++DF S PLHY DT RVYI+KSF+ +FPALR+ AI LP L+ F
 Sbjct: 242 YDVYIIEDDYLADFDSSHSLPLHYLDTDNRVYIKSFTPTLFFPALRIGALSPLNQLRDIF 301

25 Query: 302 MAYKKLMDYDTNLIQKALALYIENGLYAKNSQYLYRYQKDLANSKSLADHP-NLPSY 360
 + +K L+DYDTNLI+QKAL+LYI+NG++A+N+Q+L + Y K L + N+P Y
 Sbjct: 302 IKHKSLLIDYDTNLIQKALSLEYIDNGMFARNTQHLHHIYHAQWNKIKDCLEKYALNIP-Y 360

Query: 361 SLHEDSVLFDCK 373
 + SV F SK
 Sbjct: 361 RIPKGSVTFQLSK 373

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6885> which encodes the amino acid
 sequence <SEQ ID 6886>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3043(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 176/382 (46%), Positives = 255/382 (66%), Gaps = 7/382 (1%)

Query: 1 MVTKVEEIRSYLIASIQNGKLAPGDRLPISIRQLANQFSCNKDITVQVLMELRFDNYIYAK 60
 M TK + I S + IQ +L GD+LPSIR L+ + C+KDTVQR L+EL++ + IYA
 Sbjct: 1 MTKYQPTIISNIEQDIQKQRLKKGDKLPISIRVLSKVYCSKDTVQRALLELKYRHLIYAV 60

45 Query: 61 PRSGYVFDSSHQEEVEEGVSLPNSEIANIAYDDFRLCLNETLIGREDYLFNYYYRQEGLL 120
 P+SGYVV + + ++L + N+AY+DFRLCLNE L ++ YLF+YY++ EGL
 Sbjct: 61 PKSGYVVL-GNVSMFDNVLNLSLEDYNNMAYEDFRLCLNEALSADKYLPHYYHKTEGLE 119

50 Query: 121 DLSKAVAKLMEETGVYVPLDDIVITAGTQQALFILTQVTFPNRKSRLVIEEPTYPRMIEL 180
 +L +A+ + E VY D ++IT+GTQQAL+IL+Q+ FEN +L+E+PTY RM +
 Sbjct: 120 ELREALLLYLAENSVYSNKDQLLITSGTQQALYILSQMPFPNTGKTILLEKPTYHRMEAI 179

55 Query: 181 IKTNLPHYETISRGTGIDFQRLLEEIFQTQSIKFFYVIPRMHNPLGTSYNPVMKRLIEM 240
 + LPY+TISR +G+D + LE +FQT IKFFY I R +PLG SY+ E + ++ +
 Sbjct: 180 VAQGLPYQTISRHFNGLDLELLESIFQTGDIKFFYTISRFSHPLGLSYSTKEKEAIVRL 239

60 Query: 241 AEKYDVYIVEDDYMSDFA--SQSPLHYDTHGRVIYLSFSKAIFFPALRLAAICLPQALK 298
 A++Y VYI+EDDY+ DF + P+HYDTH R+IYLSFS ++FPALR+ A+ LP LK
 Sbjct: 240 AQRYQVYILEDDYLGFVVKLEPPHYYDTHHRIIYLSFSMSVFPALRIGALVLPSCGLK 299

65 Query: 299 STFMAKKLMDYDTNLIQKALALYIENGLYAKNSQYLYRYQKDLANSKSLADHPNLP 358
 F+ K L+D DTNL++QKALALY+ENG++ KN +++K RY K ++ N P
 Sbjct: 300 PHFLTQKSLIDLDTNLLMQKALALYLENGMFQKNLRFTK-RYLQQRERQLALFLKQ-NCP 357

-2515-

Query: 359 S--YSLHSDSVLFDCSKLDNFK 378
 Y L ++ D + D+++
 Sbjct: 358 DIHYQLTPTHLVIDYTTSDSYR 379

- 5 SEQ ID 6884 (GBS423) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 7; MW 49.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 2; MW 74kDa).

GBS423-GST was purified as shown in Figure 219, lane 2-3.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 2230

A DNA sequence (GBSx2349) was identified in *S.galactiae* <SEQ ID 6887> which encodes the amino acid sequence <SEQ ID 6888>. This protein is predicted to be pseudouridylate synthase I (truA). Analysis of this protein sequence reveals the following:

- 15 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.3265 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:BAB03886 GB:AP001507 tRNA pseudouridine synthase A
 (pseudouridylate synthase I) [Bacillus halodurans]
 Identities = 105/240 (43%), Positives = 147/240 (60%), Gaps = 2/240 (0%)
- Query: 1 MTRYKAQISYDGSFAFSGFORQPNCRVTQEEIERTLKRLNSGNDVIIHGAGRTDVGHVHAYG 60
 M R +++YDG+ F+G+Q QPN RTVQ E+E LK ++ G + + +GRTD GVHA G
 30 Sbjct: 1 MKRIGLKVAYDGTDFAGYQIQPNERTVQGELESVLKNIHKGMSIRVTASGRITDGVHARG 60
- Query: 61 QVIHFDPQARDVEKLRFGLDTCQPDIDIVKVEQVSDDFHCRYDKHIKTYEFLVDIGRP 120
 Q++HFD + V++ L++Q P DI +++ V DFH RY K Y + V
 35 Sbjct: 61 QIVHFDTSLSFPVDRWPILALNSQLPADICVLEAADVPADFHARYSAKTKEYRYRVLTSAQ 120
- Query: 121 KNPMRNYATHYPYVVIIELMQEAIKDLVGTHDFTGFTASGTSVENKVRTIFDAKIQFEA 180
 + RNY H YP+ +E MQ A L+GTHDF+ F A+ VE+KVRTI D + E
 40 Sbjct: 121 ADVFRRNYTYHVRYPDLVEAMQRAAVQLLGTHDFSSFCAAKAEVEDKVRTIEDVALWREG 180
- Query: 181 SKNLLIFTFTGNGFLYKQVRNMVGTLLKIGNRMPISQIKTILQAKNRDLAGPTAAGNGL 240
 + LIF+ GNGFLY VR +VGTLL+IG G+ ++ IL A++R+ AG TA G+GL
 45 Sbjct: 181 DE--LIFSIRGNGFLYNMVRIVGTLLLEIGAGKRSABEVAKILAARSREAAGKTAPGHGL 238

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6889> which encodes the amino acid
 45 sequence <SEQ ID 6890>. Analysis of this protein sequence reveals the following:

- Possible site: 58
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.2558 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2516-

Identities = 184/249 (73%), Positives = 214/249 (85%)

Query: 1 MTRYKAQISYDGSFSGFQRPNCRTVQEEIERTLKRLNSGNDVIIHGAGRTDVG VHAYG 60
 M RYKA ISYDG+ FSGFQRP + RTVQEEIE+TL +LN+G +IIHGAGRTD GVHAYG
 5 Sbjet: 1 MVRYSKATISYDGTLSFGFQRPRLRTVQEEIEKTLKLNNGTKIIHGAGRTDAGVHAYG 60

Query: 61 QVIHFDLPQARDVEKLRFLDTQCPDDIDIVKVEQVSDDFHCRYDKHIKTYEFLVDIGRP 120
 QVIHFDLPQ ++VEKLRFLDTQ P+DID+V +E+V+DDFHCRY KH+KTYEFLVD GRP
 10 Sbjet: 61 QVIHFDLPQEVEKLRFLDTQTPEDIDVNIKVADDFHCRYQKHLKTYEFLVDNGRP 120

Query: 121 KNPMRMNYATHYPYPVIEELMQEAIKDLVGTHDFTGFTASGTSVENKVRTIFDAKIQFEA 180
 KNPMRM+Y THYPY + I+LMQEAI LVGTHDFTGFTA+GTSV+NKVRTI A + +
 Sbjet: 121 KNPMRMHYTHYPYTLNKLMEAINGLVGTHDFTGFTAAGTSVQNKVRTITKATVSRDE 180

Query: 181 SKNLLIFTFTGNGFLYKQVRNMVGTLLKIGNRMPISQIKTILQAKNRDLAGPTAAGNGL 240
 + L+FTF+GNGFLYKQVRNMVGTLLKIGNG+MP+ Q+K IL +KNR LAGPT +GNGL
 15 Sbjet: 181 KTDFLVFTFSGNGFLYKQVRNMVGTLLKIGNGQMPVEQVKVILSSKNRQLAGPTISGNGL 240

Query: 241 YLKEIITYED 249
 YLKEI YE+
 20 Sbjet: 241 YLKEICYEN 249

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2231

A DNA sequence (GBSx2350) was identified in *S.agalactiae* <SEQ ID 6891> which encodes the amino acid sequence <SEQ ID 6892>. This protein is predicted to be phosphomethypyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2051(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15828 GB:Z99123 phosphomethylpyrimidine kinase [Bacillus subtilis]
 Identities = 95/253 (37%), Positives = 150/253 (58%), Gaps = 13/253 (5%)
 40 Query: 1 MKTRNVLAISGNDIFSGGGLHADLATVNVNKLHGFVAVTCLTAMSDKG---FEVPIEAS 57
 M L I+G+D G G+ ADL T+ ++G A+T + AM +V PI+
 Sbjet: 1 MSMHKALTIAGSDSSGGAGIQADLKTFOEKNVYGMTALTIVIVAMPNNSWNHQVFPIDTD 60

Query: 58 ILKQQLLESKD-VEFGSIKGLLPNVETAQVVLEFVKSQCECPVVLDPVLVCKENHDL-- 114
 ++ QL ++ D + ++K G+LP V+ ++ + +K KQ VV+DPV+VCK +++
 45 Sbjet: 61 TIRAQLATITDGGIGVDAMKTGMLPTVDIIELAAKTIKEKQLKNVVIDPVMVCKGANEVLY 120

Query: 115 --EVSQRLREQLIAFFPYADVITPNLVEAQLLTGLS-IENLDQMIAAEKLYDMGAKHVVI 171
 LREQL P A VITPNL EA L+G+ ++ +D M AA+K++ +GA++VVI
 50 Sbjet: 121 PEHAQLREQLA---PLATVITPNLFEASQLSGMDELKTVDMMIEAAKKIHALGAQYVVI 177

Query: 172 KGGNRLNABEATDLYDGERFETYVFPVVDANNT-GAGCTFASSIASQLAMGKNVEDAVK 230
 GG +L E+A D+ YDGE E ++D T GAGCTF++++ ++LA G V++A+
 55 Sbjet: 178 TGGGKLKHEKAVDVLVDGETAEVLESEMIDTPYTHGAGCTFSAAVTAELAKGAEVKEAIY 237

Query: 231 MSKGFVYQAIKAS 243
 +K F+ AIK S
 60 Sbjet: 238 AAKEFTTAAIKES 250

-2517-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4407> which encodes the amino acid sequence <SEQ ID 4408>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2029(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/252 (53%), Positives = 174/252 (68%)

Query: 1 MKTRNVLAISGNDIFSGGGLHADLATYVVKLHGFVAVTCLTAMSDKGFEVIPIEASILK 60

MKT ++ ISGNDI SGGGL+ADLATY+ L FVAVTCLT S++GF + P+ I +

Sbjct: 1 MKTDYIVTISGNDILSGGGLYADLATYIRYDLQAFVAVTCLTTRSEEGFSLFPVAKEIFR 60

Query: 61 QQLESKDVDFGSIKLGLLPNVETAQVVLEFVKSQCEPVLDPVLVCKENHDLEVSQLR 120

QL S + +IK+GLLPN E ++VL+F+K PVVLDPV L CKE D+++ LR

Sbjct: 61 DQLNSFTNAPISAIKIGLLPNAEMCEIVLDFIKGHLGIPVLDPLACKEDIDVKIVPLR 120

Query: 121 EQLIAFFFPVADVITPNLVEAQLLTGLSIENLDQMKAIEKLYDMGAKHVVIKGGNRLNAE 180

++++ PY V+TPNLVEAQLL+ I +L M+ AA+ Y +GAK VVIKGGNR + +

Sbjct: 121 QEILQLLPVVTVPVTPNLVEAQLLSQKEIVSLKDMQEAARYFYQLGAKQVVIKGGNRFSQK 180

Query: 181 EATDLYYDGERFETYVPVVDANNTGAGCTFASSIASQLAMKNVEDAVKMSKGFVYQAI 240

+A DL+YDG+ T PV++ NN GAGCTFASSIASQL K +AVK SK VYQAI

Sbjct: 181 KAIDLFYDGKEIVTLECPVLEKKNIGAGCTFASSIASQLVKKKTPEAVKNSKELVYQAI 240

Query: 241 KASDKYGVVQHF 252

SD+YGV Q +

Sbjct: 241 LQSDRYGVKQSY 252

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2232

A DNA sequence (GBSx2351) was identified in *S.agalactiae* <SEQ ID 6893> which encodes the amino acid sequence <SEQ ID 6894>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -6.05 Transmembrane 97 - 113 (96 - 119)

INTEGRAL Likelihood = -0.22 Transmembrane 54 - 70 (54 - 70)

----- Final Results -----

bacterial membrane --- Certainty=0.3421(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA30952 GB:AP000007 202aa long hypothetical protein [Pyrococcus horikoshii]

Identities = 48/148 (32%), Positives = 78/148 (52%), Gaps = 9/148 (6%)

Query: 10 VQLAIVTAISIVLGMFISIPTPTGFLTLDDAGIFFAIFYFGKKEGAVVGALAGFLIDLK 69

V A+VTA+++V+ I IP G+L D I + FG G G + DLL

Sbjct: 49 VMAALVTAMTMVIR--IPIPASQGYLNFQDIMIMLTSVLFGPLVGGFAGGVGSAFADLL- 105

Query: 70 GYPNWMFFSLLIHGTQGYLAGLPGR-----RRLGLISATLVMVLGYAIASGLMYGWGA 123

GYP+W F+L+I GT+G + G + + LLG + VMV+GY + ++YG

-2518-

Sbjct: 106 GYPSWALFTLVIKGTEGIIVGYFSKGEANYGKILLGTVLGGSSVMVIGYVSVAYVLYGPAG 165

Query: 124 VLPDIPGNIMQNMVGMVVGFFALNKSLE 151

+ ++ +I+Q + G+V+G L L++

5 Sbjct: 166 AIGELYNDIVQAVSGIVIGGGLGYILKK 193

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6895> which encodes the amino acid sequence <SEQ ID 6896>. Analysis of this protein sequence reveals the following:

Possible site: 54

10

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -4.62 Transmembrane 98 - 114 (97 - 119)

INTEGRAL Likelihood = -0.00 Transmembrane 135 - 151 (135 - 151)

15

----- Final Results -----

bacterial membrane --- Certainty=0.2848(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20

The protein has homology with the following sequences in the databases:

>GP:CAB49310 GB:AJ248284 hypothetical protein [Pyrococcus abyssi]

Identities = 42/145 (28%), Positives = 73/145 (49%), Gaps = 10/145 (6%)

25

Query: 7 RQMSLTGILTALVVVLGRFVMLPTPT--GFLTLLDAGIYAVSFSGSAQGAIVGGLSGFL 64

R ++++ + ALV + + +P P G+L D I V+ FG G GG+ +

Sbjct: 39 RTVAISAVAAALVTAMTMVIRIPIPASQGYLNFQDIMIMLVAVLFGPLVGGFAGGVGSAI 98

30

Query: 65 IDLVAGYFPQWMFHSLIAHSVQGYFAGWRGR-----KRWLGVVIGSFIMIFWYFLGSLML 118

DL+ GYP W +LI +G G+ + K +G V+G FIM+ Y S +L

Sbjct: 99 ADLI-GYPSWALFTLIKSGEGLVVGYPFSKGEPNYSKILIGTVLGGFIMVLGYVSVSVVL 157

Query: 119 GYGLSGSLAGIWNVMQNTLGLFVG 143

YG +G+++ ++ + +Q G+ +G

Sbjct: 158 -YGPAGAISELYNDTVQAVSGIVIG 181

35

An alignment of the GAS and GBS proteins is shown below.

Identities = 77/155 (49%), Positives = 106/155 (67%), Gaps = 1/155 (0%)

40

Query: 1 MRKEKTSQLVQLAIVTAISIVLGMFISIPTPTGFLTLLDAGIFFAIFYGKKEGAVVGAL 60

M+ K Q+ I+TA+ +VLG F+ +PTPTGFLTLLDAGI+ +F FG +GA+VG L

Sbjct: 1 MQNSKIRQMSLTGILTALVVVLGRFVMLPTPTGFLTLLDAGIYAVSFSGSAQGAIVGGL 60

45

Query: 61 AGFLIDLKGYPNWMMFFSLIHGTQGYLAGLPGRRLRLGLISATLVMVLGYAIASGLM-Y 119

+GFLIDL+ GYP WMF SL+ H QGY AG GR+R LG++ + +M+ Y + S ++ Y

Sbjct: 61 SGFLIDLVAGYFPQWMFHSLIAHSVQGYFAGWRGRKRWLGVVIGSFIMIFWYFLGSLMLGY 120

Query: 120 GWGAVLPDIPGNIMQNMVGMVVGFFALNKSLEK 154

G L I GN+MQN +G+ VGF + K++ R KK

Sbjct: 121 GLSGSLAGIWNVMQNTLGLFVGFIIFKAILRQKK 155

50

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2233

55 A DNA sequence (GBSx2352) was identified in *S.galactiae* <SEQ ID 6897> which encodes the amino acid sequence <SEQ ID 6898>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2519-

bacterial cytoplasm --- Certainty=0.0881(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15708 GB:Z99122 alternate gene name: ipc-33d [Bacillus subtilis]
 Identities = 91/176 (51%), Positives = 115/176 (64%)

10 Query: 6 NKLKQETKAIVVDIERSALKKGQIFVLGLSSSEVSGGLIGNSSSEIGEIIIVEVILKEL 65
 N+LKQ K ++ + +++ LK+ Q+FVLG S+SEV+G IG + S +I E I + +
 Sbjct: 2 NELKQTKWTKMLSEFQDQAEKQDQLFVLGCSTSEVAGSRIGTSGSVDIAESIYSGLAELR 61

15 Query: 66 HSRGIYLAVQGCCEHVNRLVVEAELAEERQQLEVVNVVPNLHAGGSGQVAAFKLMTSPVEV 125
 GI+LA Q CEH+NRALVVEAE A+ +L V+ VP AGG+ AFK M SPV V
 Sbjct: 62 EKTGIHLAFQCCEHLNRLVVEAETAKLFRILPTVSAVPVPKAGGAMASYAFKQMKSPVLV 121

20 Query: 126 EEIVAHAGIDIGDTSIGMHKRVQVPLIPISRELGGAHVTALASRPKLIGGARAGY 181
 E I A AGIDIGDT IGMH+K V VP+ LG AHVT +RPKLIGG RA Y
 Sbjct: 122 ETIQADAGIDIGDTFIGMHLKPVAVPVRVVSQNSLGSAAHVTIARTRPKLIGGVRAVY 177

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6899> which encodes the amino acid sequence <SEQ ID 6900>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2166(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 132/183 (72%), Positives = 161/183 (87%)

35 Query: 6 NKLKQETKAIVVDIERSALKKGQIFVLGLSSSEVSGGLIGNSSSEIGEIIIVEVILKEL 65
 N L+++T+ IV+D++ERSA++ G +FVLGLSSSE+ G IGK SS E+G+I+VEV+L EL
 Sbjct: 3 NNLEKQTRREIVIDVVERSAIQPGNLFVLGLSSSEILGSRIGKQSSLEVGGIIVVEVVLDEL 62

40 Query: 66 HSRGIYLAVQGCCEHVNRLVVEAELAEERQQLEVVNVVPNLHAGGSGQVAAFKLMTSPVEV 125
 + RG++LAVQGCCEHVNRLVVE +AE +QLE+VNVVPNLHAGGS Q+AAF+LM+ PVEV
 Sbjct: 63 NKRGVHLAVQGCCEHVNRLVVERHVAESKQLEIVNVVPNLHAGGSAQMAAFQLMSDFVEV 122

45 Query: 126 EEIVAHAGIDIGDTSIGMHKRVQVPLIPISRELGGAHVTALASRPKLIGGARAGYTS DP 185
 EE++AHAG+DIGDT+IGMHKRVQ+PLIP RELGGAHVTALASRPKLIGGARA Y D
 Sbjct: 123 EEVIAHAGLDIGDTAIGMHKRVQIPLIPCQRELGGAHVTALASRPKLIGGARADYNMDI 182

Query: 186 IRK 188
 IRK
 Sbjct: 183 IRK 185

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2234

A DNA sequence (GBSx2353) was identified in *S.agalactiae* <SEQ ID 6901> which encodes the amino acid sequence <SEQ ID 6902>. Analysis of this protein sequence reveals the following:

55 Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -11.25 Transmembrane 21 - 37 (13 - 46)
 INTEGRAL Likelihood = -4.30 Transmembrane 78 - 94 (76 - 113)
 INTEGRAL Likelihood = -2.07 Transmembrane 96 - 112 (95 - 113)

-2520-

----- Final Results -----

5 bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 105/261 (40%), Positives = 150/261 (57%), Gaps = 2/261 (0%)

10 Query: 12 NVEEVLFTHFFTKLIS--ILLIIIAFVIVRQVINYLFEKTVNRSIAFSRQKVARQKTLAKL 69
 N+ F T +I+ +L+ +IAF+IVR + + + R ++ R TL KL
 Sbjct: 7 NITSGAFLASTFIIAGKVLVAVIAFLIVRAIGKRIISNSFARMAKNNQLSSGRVVTLEKL 66

15 Query: 70 SHNVLYNTLYFFLFYWILSVISLLAGAGIAGVAIGLGAQGFLSDVVGFFILLEN 129
 S N +YTL F +L+I G+ S+L+AGAGI G+AIG GAQG +SD+V GFFILLE
 Sbjct: 67 SLNAFSYTLMFIFATTLTIFGLNPSALIAGAGIVGLAIGFGAQLVSDIVTGFFILLEK 126

20 Query: 130 QFDVGDIIINVGTVSGTVTNVGIRTQIHDFDGLTHFIPNRNITIVSNKSRNMRAQIDIP 189
 Q DVG D + G V G V VG+RT I FDGTLH+IPNRNI VSN SR NMRA +DI
 Sbjct: 127 QIDVG DYVTAGGVDGIVEEVGLRTALIRGFDGTLHYIPNRNIANVSNHSGNMALVDIS 186

25 Query: 190 LRVHTNLDQISDIVTKINEEYVSKHPAIVGEPTVFGPTTNANGQFVYRINIFTQNGAQFD 249
 + + N+D+ ++ K+ ++ + I+ P V G + V RI T+N Q+
 Sbjct: 187 ISYNDNIDEAISVMQKVCQDLAEQDERIIEGPDVIGVQNLGDSVIVIRIIAKTENMEQWS 246

Query: 250 IYAEFYKLYQKALEEGDILP 270
 + K ++A+ I++P
 Sbjct: 247 VERLLRKQLKEALEAHNIEIP 267

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6903> which encodes the amino acid sequence <SEQ ID 6904>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

35 INTEGRAL Likelihood = -8.49 Transmembrane 24 - 40 (15 - 45)
 INTEGRAL Likelihood = -4.83 Transmembrane 78 - 94 (73 - 99)
 INTEGRAL Likelihood = -2.07 Transmembrane 96 - 112 (95 - 113)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.4397(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:BAB06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 104/249 (41%), Positives = 151/249 (59%), Gaps = 4/249 (1%)

50 Query: 22 KKLVSLLIILLFFAILKRVNTNYLFEKTIKSFAYSRSQSEARKKTLTKLTHNINILNYLLYFL 81
 K LV++I L+ AI KR+ + F + + + S R TL KL+ N +Y L F+
 Sbjct: 23 KVLVAVIAFLIVRAIGKRIISNSFARMAKNN---QLSSGRVVTLEKLSLNAFSYTLMFI 78

55 Query: 82 LIYWILSLFGIFVSSLLAGAGIAGVAIGLGAQGFLSDVVGFFILFENQFEVGDNVITISD 141
 +L++FG+ S+L+AGAGI G+AIG GAQG +SD+V GFFIL E Q +VGD VT
 Sbjct: 79 FATTLTIFGLNPSALIAGAGIVGLAIGFGAQLVSDIVTGFFILLEKQIDVGDYVTAGG 138

60 Query: 142 IEGSVFVGIVGIRTQIRGFDGTLHFIPIRNSITVSNKSRGNMRALIEIPLYSTVNLSQVTR 201
 ++G V VG+RT IRGFDGTLH+IPNR+I VSN SRGNMRAL++I + N+ +
 Sbjct: 139 VDGIVEEVGLRTALIRGFDGTLHYIPNRNIANVSNHSGNMALVDISYNDNIDEAIS 198

Query: 202 IIDEVNQKELPNHPQIVGKPNILGFQNNNSNGQFTFRIAIFTENGEQFKIYHTFYRLYQEA 261
 ++ +V + +I+ P+++G QN + RI TEN EQ+ + + +EA
 Sbjct: 199 VMQKVCQDLAEQDERIIEGPDVIGVQNLGDSVIVIRIIAKTENMEQWSVERLLRKQLKEA 258

Query: 262 LLKEGIQLP 270

-2521-

L I++P
 Sbjct: 259 LEAHNIEIP 267

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 164/265 (61%), Positives = 215/265 (80%)

Query: 7 FIDHLNVEEVLFTEFTKLISILLIIAFVIVRQVINYLFEKTVNRSIAFSRQKVARQKTL 66
 +++ ++E + T F KL+S+++L++ F I+++V NYLFEKT+N+S A+SRQ AR+KTL
 10 Sbjct: 7 YLEQSHIENIGLTIFKKLVSLIILLFFAILKRVTNYLFEKTINKSFAYSRSQSEARKKTL 66

Query: 67 AKLSHNVLNNTLYFFLFYWILSILGVPISLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL 126
 +KL+HN+LNY LYF L YWILS+ G+P+SSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL
 15 Sbjct: 67 SKLTHNILLNLLYFLLIYWILSLFGIPVSSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL 126

Query: 127 LENQFDVGDIIINVGTVSGTVTNVGIRTTQIHDFDGTLHFIPNRNITIVSNKSRNMRAQI 186
 ENQF+VGD + + + G+V VGIRTTQI FDGTLHFIPNR+IT+VSNKSR NMRA I
 20 Sbjct: 127 FENQFEVGDNVTISDIEGSGVFGVIRTTQIRGFDGTLHFIPNRSTTVVSNKSRNMRAI 186

Query: 187 DIPLFVHTNLDQISDIVTKINEEYVSKHPAIVGEPTVFGPTTNANGQFVYRINIFTQNGA 246
 +IPL+ NL Q++ I+ ++N++ + HP IVG+P + GP N+NGQF +RI IFT+NG
 25 Sbjct: 187 EIPLYSTVNLSQVTRIIDEVNQKELPNHPQIVGKPNILGPONNSNGQFTFRIAIFTENG 246

Query: 247 QFDIYAEFYKLYQKAILEEGIDLPT 271
 QF IY FY+LYQ+A+L+EGI LPT
 30 Sbjct: 247 QFKIYHTFYRLYQEALLKEGIQLPT 271

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2235

30 A DNA sequence (GBSx2354) was identified in *S.agalactiae* <SEQ ID 6905> which encodes the amino acid sequence <SEQ ID 6906>. This protein is predicted to be RopA (tig). Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1785(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40

A related GBS nucleic acid sequence <SEQ ID 9283> which encodes amino acid sequence <SEQ ID 9284> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6907> which encodes the amino acid sequence <SEQ ID 6908>. Analysis of this protein sequence reveals the following:

45 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0776(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

55 Identities = 303/354 (85%), Positives = 337/354 (94%)

Query: 1 MSTSFENKATNNGIITFTISQDEIKPALDQAFNKVKKDLNVPGFRKGHPRTVFNQKFGE 60

-2522-

MSTSFENKATNRG+ITFTISQD+IKPALD+AFNK+KKDLN PGFRKGHMPR VFNQKFGE
 Sbjet: 30 MSTSFENKATNRGVITFTISQDKIKPALDKAFNKIKKDLNAPGFRKGHMPRPVFNQKFGE 89
 Query: 61 EALYENALNLVLPKAYEAAVAELGLDVVAQPKIDVVSMEKGQDWKLTAEVVTKEPVKLG 120
 5 E LYE+ALN+VLP+AYEAAV ELGLDVVAQPKIDVVSMEKG++W L+AEVVTKEPVKLG
 Sbjet: 90 EVLYEDALNIVLPEAYEAAVTELGLDVVAQPKIDVVSMEKGKWTLSAEVVTKEPVKLG 149
 Query: 121 YKDLSEVDASKEVSDEEVDKVERERNLAEELTVKDGEAAQGDVVDFVGSVDGVEFD 180
 YK+L VEVDASKEVSDE+VDAK+ERER NLAEEL +KDGEAAQGDVVDFVGSVDGVEFD
 10 Sbjet: 150 YKNLVSEVDASKEVSDEEDVDKIERERQNLAEELIKDGEAAQGDVVDFVGSVDGVEFD 209
 Query: 181 GGKGNFSLELGGSGQFIPGFEEQLVGSKAGQTVDNVTFPEDYQAEDLAGKDAKFVTTIH 240
 GGKGNFSLELGGSGQFIPGF+QLVG+KAG V+VNVTFPE YQAEDLAGK AKF+TTIH
 15 Sbjet: 210 GGKGNFSLELGGSGQFIPGFEDQLVGAKAGDEVEVNVTFPESYQAEDLAGKAAKFMTTIIH 269
 Query: 241 EVKTKEVPALDDELAKDIDDEVEITLDELKAKYRKELESAKEIAFDDAVEGAAIELAVANA 300
 EVKTKEVP LDDELAKDID++V+TL++LK KYRKELE+A+E A+DDAVEGAAIELAVANA
 Sbjet: 270 EVKTKEVPELDDELAKDIDEDVDITLEDLKVYRKELEAAQETAYDDAVEGAAIELAVANA 329
 20 Query: 301 EIVELPEEMVHDEVHRAMNEFMGNMQRQGISPEMYFQLTGTTTEEDLHKQYQADA 354
 EIV+LPEEM+H+EV+R++NEFMGNMQRQGISPEMYFQLTGTT+EDLH QY A+A
 Sbjet: 330 EIVDLPEEMIHEEVNRSVNEFMGNMQRQGISPEMYFQLTGTTQEDLHNQYSAEA 383

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 2236

A DNA sequence (GBSx2355) was identified in *S.agalactiae* <SEQ ID 6909> which encodes the amino
 acid sequence <SEQ ID 6910>. This protein is predicted to be galactose-6-phosphate isomerase laca subunit
 (rpiB). Analysis of this protein sequence reveals the following:

30 Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.3491(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAA25177 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
 Identities = 92/141 (65%), Positives = 115/141 (81%)
 Query: 1 MTIIIGADAHGVELKEVIRQHLTSLGKEIIDLTDTSKDFVDNTLAIKVNQKEDNLGIM 60
 M I++GAD G LK+V++ L G E+ID+T +DFVD TLA+ ++VN+ E NLGI+
 45 Sbjet: 1 MAIVVGADLKGTRLKDVVKNFLVEEGFEVIDVTKDQDFVDVTLAVASEVKNKDEQNLGIV 60
 Query: 61 VDAYGVGPFPMVATKVKGMIAAEVSDERSAYMTRAHNNARMITLGSEIVGPGVAKHIVEGF 120
 +DAYG GPFPMVATK+KGM+AAEVSDERSAYMTR HNNARMIT+G+ETVG +AK+I + F
 Sbjet: 61 IDAYGAGPFPMVATKIKGMVAAEVSDERSAYMTRGHNNARMITVGAIEVGDELAKNIAKAF 120
 50 Query: 121 VDGTYDAGRHRQIRVDMLNKMC 141
 V+G YD GRHQ+RVDMLNKMC
 Sbjet: 121 VNGKYDGGRRHQVRVDMLNKMC 141

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6911> which encodes the amino acid
 55 sequence <SEQ ID 6912>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

-2523-

bacterial cytoplasm --- Certainty=0.3224 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 101/140 (72%), Positives = 117/140 (83%)

Query: 1 MIIIGADAHGVELKEVIRQHLTSLGKEIIDLTDTSKDFVDNTLAIVAKVNQKEDNLGIM 60
 M II+GADAHG LKE+I+ L G +IID+TD + DF+DNTLA+ VN+ E LGIM
 10 Sbjct: 1 MAILIGADAHGNALKEIKSFLQEEGYDIIDVTDINSDFDINTLAVAKAVNEAEGRLGIM 60

Query: 61 VDAYGVGPFMVATKVKGMIAAEVSDERSAYMTRAHNNARMITLGSEIVGPGVAKHIVEGF 120
 VDAYG GPFMVATK+KGM+AAEVSDERSAYMTR HNNARMIT+G+EIVGP +AK+IV+GF
 15 Sbjct: 61 VDAYGAGPFMVATKLGKGMVAAEVSDERSAYMTRGHNNARMITIGAEIVGPELAKNIVKGF 120

Query: 121 VDGTYDAGRHRQIRVDMLNKM 140
 V G YD GRHQIRVDMLNKM
 Sbjct: 121 VTGPYDGRHRQIRVDMLNKM 140

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2237

A DNA sequence (GBSx2356) was identified in *S.agalactiae* <SEQ ID 6913> which encodes the amino acid sequence <SEQ ID 6914>. This protein is predicted to be galactose-6-phosphate isomerase lach subunit (rpiB). Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.2511 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10189> which encodes amino acid sequence <SEQ ID 10190> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA25178 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
 Identities = 138/171 (80%), Positives = 157/171 (91%)

40 Query: 10 MKIAVGCDHIVTYDKIAVVDYLKTKGYEVIDCGTYDNI RTHYPIYGKKVGEAVASGKADL 69
 M+IA+GCDHIVT K+AV ++LK+KGYEV+D GTYD++RTHYPIYGKKVGEAV SG+ADL
 Sbjct: 1 MRIAIGCDHIVTDVKMAVSEFLKSKGYEVLDFGTYDHVRTHYPIYGKKVGEAVVSGQADL 60

45 Query: 70 GVCICGTGVGINNAVNKVP GIRSALVRDLTSAIYAKEELNANVIGFGGKITGGLLMTDII 129
 GVCICGTGVGINNAVNKVP G+RSALVRD+TSA+YAKEELNANVIGFGG ITGGLL DII
 Sbjct: 61 GVCICGTGVGINNAVNKVP GRSALVRDMTSALYAKEELNANVIGFGGMITGGLLMDII 120

Query: 130 EAFIRAKYKPTKENKVLIEKIAEVETHNAHQEENDFFTEFLDKWNRGEYHD 180
 EAFI A+YKPT+ENK LI KI VETHNAHQ + +FFTEFL+KW+RGEYHD
 50 Sbjct: 121 EAFIEAEYKPTENKKLIAKIEHVETHNAHQADEEFFTEFLEKWD RGEYHD 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6915> which encodes the amino acid sequence <SEQ ID 6916>. Analysis of this protein sequence reveals the following:

Possible site: 57
 55 >>> Seems to have no N-terminal signal sequence

-2524-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3048(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 136/171 (79%), Positives = 160/171 (93%)

10

Query: 10 MKIAVGC DHIVTYDKIAVVDYLKTKGYEVIDCGTYDNIRTHYPIYGKKVGEAVASGKADL 69
 MKIA+GCDHIVT +K+AV D+LK+KGY+VIDCGTYD+ RTHYPI+GKKVGEAV +G+ADL
 Sbjct: 2 MKIAIGCDHIVTNEKMAVSDFLKSGYDVIDCGTYDHTRTHYPIFGKKVGEAVVNGQADL 61

15

Query: 70 GVCICGTG VGINNAVNVKPGIRSALVRDLTSAIYAKEELNANVIGFGGKITGGLLMTDII 129
 GVCICGTG VGINNAVNVKPGIRSALVRD+T+A+YAKEELNANVIGFGGKITG LLM DII
 Sbjct: 62 GVCICGTG VGINNAVNVKPGIRSALVRDMMTALYAKEELNANVIGFGGKITGELLMCDII 121

20

Query: 130 EAFIRAKYKPTKENKVLIEKIAEVETHNAHQEENDFFTEFLDKWNRGEYHD 180
 +AFI+A+YK T+ENK LI KIA +E+H+A+QE+ DFFTEFL+KW+RGEYHD
 Sbjct: 122 DAFIKAEYKETEENKKLIAKIAHLESHHANQEDPDFTEFLKWDNRGEYHD 172

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2238

A DNA sequence (GBSx2357) was identified in *S.agalactiae* <SEQ ID 6917> which encodes the amino acid sequence <SEQ ID 6918>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq

30

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10187> which encodes amino acid sequence <SEQ ID 10188> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA25179 GB:M60447 tagatose 6-P kinase [Lactococcus lactis]
 Identities = 192/310 (61%), Positives = 236/310 (75%)

40

Query: 11 MILITVTLNPSIDISYCLENFNMDTVNRVTDVSKTPGGKGLNVTRVLSQLGDNVVATGLLG 70
 MILITVTLNPS+DISY LE +DTVNRV DVSKT GGKGLNVTRVL + GD V ATG LG
 Sbjct: 1 MILITVTLNPSVDISYPLETLKIDTVNRVKDVSKTAGGKGLNVTRVLYESGDKVTATGFLG 60

45

Query: 71 GDFGDFIRSGLEDALEIRHQFLSIGGETRHCIAVLHEGQQTEILEKGPHITKDEADAFLNH 130
 G G+FI S L+ + F I G TR+CIA+LHEG QTEILE+GP I+ +EA+ FL+H
 Sbjct: 61 GKIGEFIESELEQSPVSPAFYKISGNTRNCIAILHEGNQTEILEQGPTISHEEAEGFLDH 120

50

Query: 131 LKLIFDAATITVSGSLPKGLPSDYARLISLANHFNKKVVLDCSGEALRSVLKSSAKPT 190
 + + ++T+SGSLP CLP+DYY +LI LA+ VVLDCSG L +VLKSSAKPT
 Sbjct: 121 YSNLIKQSEVVTISGSLPSGLPNDYYEKLQIASDEGVAVVLDCSGAPLETVLKSSAKPT 180

55

Query: 191 VIKPNLEELTQLIGKPIISYSLDELKSTLQDLFRGIDWVIVSLGARGAFKHNHYYQVT 250
 IKPN EEL+QL+GK ++ ++ELK L++ LF GI+W++VSLG GAFKHK+ +Y+V
 Sbjct: 181 AIKENNEELSQLLGKEVTKDIEELKDVLESLSFGIEWIVVSLGRNGAFKHKGDVFKVD 240

Query: 251 IPKIEVINPVGSGDATVAGIASALEHQDLDDTNLLKRANVLGMLNAQETLTGHINLTYQ 310
 IP I V+NPVSGSD+TVAGIASAL + D +LLK A LGMLNAQET+TGH+N+T Y+
 Sbjct: 241 IPDIPVNPVSGSDSTVAGIASALNSKKSDADLLKHAMTLGMLNAQETMTGHVNMNTNYET 300

-2525-

Query: 311 LISQIQVKEV 320
 L SQI VKEV
 Sbjct: 301 LNSQIGVKEV 310

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6919> which encodes the amino acid sequence <SEQ ID 6920>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1178(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 184/310 (59%), Positives = 232/310 (74%), Gaps = 1/310 (0%)

Query: 11 MILFTVLNPSIDISYCLENFNMDTVNVRVTDVSKTPGGKGLNVTRVLSQLGDNVAVATGLLG 70
 +ILFTVLNP+ID+SY L+ DTVNRV DV+KTPGGKGLNV+RVL++ G+ V ATG +G
 20 Sbjct: 1 VILFTVLNPAIDVSYPLDELKCDTVNVRVDVTKTPGGKGLNVSRLNEFGETVKATGCVG 60

Query: 71 GDFGDFIRSGLEDALEIRHQFLSIGGETRHCIAVLHEGQQTEILEKGPHITKDEADAFINLH 130
 G+ GDFI + L I +F I G+TR CIA+LHEG QTEILEKGP ++ DE D F +H
 25 Sbjct: 61 GESGDFIINHLPD-SILSRFYKISGDTRTCTAILHEGNQTEILEKGPMLSVDEIDGFTHH 119

Query: 131 LKLIFDAATITVSGSLPKGLPSDYYARLISLANHFNKKVVLDCSGEALRSVLKSSAKPT 190
 K + + ++T+SGSLP G+P DYY +LI +AN KK VLDCSG AL +VLK +KPT
 30 Sbjct: 120 FKYLLNDVDVVTLSGSLPAGMPDDYYQKLIKIANLNGKKTVLDCSGNALEAVLKGDSKPT 179

Query: 191 VIKPNLEELTQLIGKPISSYSLDELKSTLQQDLFRGIDWVIVSLGARGAFKAGHGHYYQVT 250
 VIKPNLEEL+QL+GK ++ D LK LQ +LF GI+W+IVSLGA G FAKH + +Y V
 35 Sbjct: 180 VIKPNLEELSQLLGKEMTKDFDALKEVLQDELFDGIEWIIVSLGADGVFAKHKDTFYNV 239

Query: 251 IPKIEVINPVGSGDATVAGIASALEHQDDTNLLKRANVLGMLNAQETLTGHINLTYQBE 310
 IPKI++++ VGSGD+TVAGIAS L + DD LL +ANVLGMLNAQE TGH+N+ Y +
 40 Sbjct: 240 IPKIKIVSAVGSGDSTVAGIASGLANDEDDRALTKANVLGMLNAQEKTTGHVNMANYDK 299

Query: 311 LISQIQVKEV 320
 L I+VKEV
 40 Sbjct: 300 LYQSIKVKEV 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2239

- 45 A DNA sequence (GBSx2358) was identified in *S.agalactiae* <SEQ ID 6921> which encodes the amino acid sequence <SEQ ID 6922>. This protein is predicted to be tagatose 1,6-diphosphate aldolase. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0369(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA25180 GB:M60447 tagatose 1,6-diP aldolase [Lactococcus
 lactis]

-2526-

Identities = 253/325 (77%), Positives = 295/325 (89%)

Query: 1 MGLTEQKQKHMEQLSDKNGIISALAFDQRGALKRLMAKYQSEPTVSQIEALKVLVAEEL 60
 M LTEQK+K +E+LSDKNG ISALAFDQRGALKRLMA+YQ EPTV+Q+E LKVLVA+EL
 5 Sbjet: 1 MVLTEQKRKSLKLSKNGFISALAFDQRGALKRLMAQYQDTEPTVAQMEELKVLVADEL 60

Query: 61 TFYASSMLLDPEYGLPATKVLDDNAGLLLAYEKTGYDTSSTKRLPDCLDIWSAKRIKEEG 120
 T YASSMLLDPEYGLPATK LD AGLLLA+EKTGYDTSSTKRLPDCLD+WSAKRIKE+G
 10 Sbjet: 61 TKYASSMLLDPEYGLPATKALDKAAGLLLAFEKTGYDTSSTKRLPDCLDVWSAKRIKEQG 120

Query: 121 ADAVKFLLYYDVSDEVNEEKEAYIERIGSECVAEIDIPFFLEILSYDEKITDSSGIEYA 180
 ADAVKFLLYYDVSDE+N++K+AYIER+GSECVAEIDIPFFLEIL+YDE+I+D+ +EYA
 15 Sbjet: 121 ADAVKFLLYYDVSDELNQKQAYIERVGSSECVAEIDIPFFLEILAYDEEISDAGSVEYA 180

Query: 181 KIKPRKVIAMKVFSNPRFNIDVLKVEVPVNM DYVEGFAQGETAYNKATAAAYFREQDQA 240
 K+KPRKVIAMKVFS+PRFNIDVLKVEVPVN+ YVEGFA GE Y+KA AA +F+ Q++A
 20 Sbjet: 181 KVKPRKVIAMKVFS DPRFNIDVLKVEVPVNVKYVEGFADGEVVS KA EAADFFK AQEEA 240

Query: 241 TLLPYIFLSAGVPAQLFQETLVFAKEAGAKFNGVLCGRATWAGSVKEYVEKGEAGARQWL 300
 T LPYI+LSAGV A+LFQETL FA ++GAKFNGVLCGRATWAGSV+ Y+++GE AR+WL
 25 Sbjet: 241 TNLPIYLSAGVSAKLFQETLVFAHDSGAKFNGVLCGRATWAGSVEPYIKEGEKAAREWL 300

Query: 301 RTIGFQNIDELNKLQKTATSWKER 325
 RT GF+NIDELNK+L KTA+ W ++
 30 Sbjet: 301 RTTGFFENIDELNKVLVKTASPWTDK 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6923> which encodes the amino acid sequence <SEQ ID 6924>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0600(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 230/323 (71%), Positives = 276/323 (85%), Gaps = 1/323 (0%)

Query: 3 LTEQKQKHMEQLSDKNGIISALAFDQRGALKRLMAKYQSEPTVSQIEALKVLVAEELTP 62
 LTE K+K ME+LS +G+ISALAFDQRGALKR+MA++Q++EPTV QIE LK LV+EELTP
 40 Sbjet: 5 LTENKRKSMEKLS -VDGVISALAFDQRGALKRMAQHQTKEPTVEQIEELKSIVSEELTP 63

Query: 63 YASSMLLDPEYGLPATKVLDDNAGLLLAYEKTGYDTSSTKRLPDCLDIWSAKRIKEEGAD 122
 +ASS+LLDPEYGLPA++V + AGLLLAYEKTGYD ++T RLPDCLD+WSAKRIKE GA+
 45 Sbjet: 64 FASSILLDPEYGLPASRVRSSEAGLLLAYEKTGYDATTTSLRPDCLDVWSAKRIKEAGAE 123

Query: 123 AVKFLLYYDVSDEVNEEKEAYIERIGSECVAEIDIPFFLEILSYDEKITDSSGIEYAKI 182
 AVKFLLYYD+D +VNE+K+AYIERIGSEC AEDIPF+LEIL+YDEKI D++ E+AK+
 50 Sbjet: 124 AVKFLLYYDIDGDQDVNEQKAYIERIGSECAEDIPFYLEILTYDEKIADNASPEFAKV 183

Query: 183 KPRKVIAMKVFSNPRFNIDVLKVEVPVNM DYVEGFAQGETAYNKATAAAYFREQDQATL 242
 K KV EAMKVFS RF +DVLKVEVPVNM +VEGFA GE + K AA FR+Q+ +T
 55 Sbjet: 184 KAHKVNEMAMKVFSKERFGVDVLKVEVPVNMKFVEGFADGEVLFTKEEAAQAFRDQEASTD 243

Query: 243 LPYIFLSAGVPAQLFQETLVFAKEAGAKFNGVLCGRATWAGSVKEYVEKGEAGARQWLRT 302
 LPYI+LSAGV A+LFQ+TLVFA E+GAKFNGVLCGRATWAGSVK Y+E+G AR+WLRT
 60 Sbjet: 244 LPYIYLSAGVSAKLFQDTLVFAESGAKFNGVLCGRATWAGSVKVYIEEGPQAAREWLRT 303

Query: 303 IGFQNIDELNKLQKTATSWKER 325
 GF+NIDELNK+L KTA+ W E+
 65 Sbjet: 304 EGFKNIDELNKVLDKTASPWTEK 326

-2527-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2240

A DNA sequence (GBSx2359) was identified in *S.agalactiae* <SEQ ID 6925> which encodes the amino acid sequence <SEQ ID 6926>. This protein is predicted to be lacx protein, chromosomal. Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0643(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10185> which encodes amino acid sequence <SEQ ID 10186> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA25184 GB:M60447 ORF [Lactococcus lactis]
Identities = 173/298 (58%), Positives = 219/298 (73%)

Query: 24 MAITIQNHLEQVTLKALGATMTSITDSQGV EYLWQGDATYWGGQAPILFPICGSRNDCV 83
M I ++N L V K L G +TSI D G+EYLWQ D YW GQAPILFPICGS+RND
Sbjct: 1 MTIELKNEYLT VQFKTLGGQLTSIKDKDGLEYLWQADPEYWNGQAPILFPICGSLRNDWA 60

Query: 84 IYRPAQAPHFTGIIPRHGFVRHKTFDYDYSISVSRFTIKSSKEMLINYPYRFSLEITYT 143
IYRP + P FTG+I RHGFVR + F + +++SV F+IK + EML NY Y+F L + YT
Sbjct: 61 IYRPQERPFFTGLIRRHGFVRKEEFTLEEVNENSVTFSIKPNAEMLDNYLYQFELRVVYT 120

Query: 144 LRNKSIAITYIVKNLESEKNMPYAIGAHPGFNCPLFEKEVFSDDYILEFEQFETCTIPESF 203
L KSI + V NLE+EK MPY IGAHP FNCPL E E + DY LEF + E+C+IP+SF
Sbjct: 121 LNGKSIRTEFQVTNLETEKTMPIYFIGAHPAFNCPLVEGEKYEDYSLEFSEVESCSIPKSF 180

Query: 204 PDTGLLDLQARHPFLENQKQLSLNHALFEKDAITLDQLRSKTVYVKSRNHAKGIQLDFDD 263
P+TGLLDLQ R PLENQK L L+++LF DAITLD+L+S++V L+SR KG+++DFDD
Sbjct: 181 PETGLLDLQDRTPFLENQKSLDLDSLFSHDAITLDRLKRSVTLRSRKSGKGLRVDFDD 240

Query: 264 FENLILWTSNNGGPFLEPWSLSLSTSI ESDILEDKQNI VRLNPKQSKQHSIRITIL 321
F NLILW++ N PF+ALEPWS LSTS+EE +ILEDK + ++ P + + S ITIL
Sbjct: 241 FPNLILWSTINKSPFIALEPWSGLSTSL EEGNILEDKPQVTKVLPDTSKKSVDITIL 298

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2241

A DNA sequence (GBSx2361) was identified in *S.agalactiae* <SEQ ID 6927> which encodes the amino acid sequence <SEQ ID 6928>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3272(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2528-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10183> which encodes amino acid sequence <SEQ ID 10184> was also identified.

- 5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA51350 GB:X72832 leucine rich protein [Streptococcus
equisimilis]
Identities = 101/278 (36%), Positives = 160/278 (57%), Gaps = 1/278 (0%)

10 Query: 10 MDFKELFPEVITKQEVKQSEDIYIIVEQDGHVLHFPKSSLTKRELYLLQMTPSLEDASSVD 69
M+ K+ FPE+ +++++ V++ +HFPKS L+++E LL++ +
Sbjct: 1 MELKDYFPFEMQVGPPLGDKWVSVKEGDQYVHFPKSLSEKERLLLEVGLGQYEVLP-P 59

15 Query: 70 SQNPWYRYLVEGRGRLPQSHSAVQFIFIEHQFTLSEELKDFLSPLVINVETIMTINQTS 129
+PW RYL++ +G PQ QFI++ HQ L +L + L ++ +E I+ I+ TQ+
Sbjct: 60 LGSPWQRYLLDHQGNPPQLFETSQFIYLNHQQVLPADLVELLQQMIAGLEVILPISTTQT 119

20 Query: 130 VMILNQDNFFNATELLTDILPTIENDFNTRLRCYFGNSWTHLQAVDWKELYEEYKLFITL 189
+ Q L +LPT+E+DF L + GN+W + A +E +EEE +L T
Sbjct: 120 AFLCRQATSIVLRSLEGLLPLESDFGALATMFVGNAWYQVAAGTLRECFEEECQLLTA 179

25 Query: 190 FLSHKAEQHYCRFPKMAWLANQSPMPSIAKAKCLQHILDTSDTSAIIKALWQEQGNLAK 249
+L K+ F ++ LW++ + P++ + Q + SD + ++ ALW E GNL +
Sbjct: 180 YLKQKSGGKLLTFAEVMWLSILSHQSFPAITRQFHQFLNPQSDMADVVHALWSEHGNLVQ 239

Query: 250 TAKALFIHRNSLQYKLDKFTQSSGLNLKILDDLAYAYL 287
TA+ L+IHRNSLQYKLDKF Q SGL+LK LDDLA+AYL
Sbjct: 240 TAQRLYIHRNSLQYKLDKFAQQSGLHLKQLDDLAFAYL 277

- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6929> which encodes the amino acid sequence <SEQ ID 6930>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

- 35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4332(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 40 An alignment of the GAS and GBS proteins is shown below.

Identities = 106/287 (36%), Positives = 169/287 (57%), Gaps = 4/287 (1%)

45 Query: 3 KTVVED-AMDFKELFPEVITKQEVKQSEDIYIIVEQDGHVLHFPKSSLTKRELYLLQM-TP 60
KTV++ AM+ K+ FPE+ +D++ +++ +HFPKS L+++E LL++
Sbjct: 7 KTVMKGMAMELKDYFPFEMQVGPPLGDKDWMSIKEGDQYVHFPKSLSEKERLLLEVGLG 66

50 Query: 61 SLEDASSVDSQNPWYRYLVEGRGRLPQSHSAVQFIFIEHQFTLSEELKDFLSPLVINVET 120
E + S PW RYL++ +G PQ + QFI++ HQ L ++L + L ++ +E
Sbjct: 67 QCEVLQPLGS--PWQRYLLDHQGNPPQLYETSQFIYLNHQQALPDDLVELLQQMIAGLEV 124

55 Query: 121 IMTINQTSQVMILNQDNFFNATELLTDILPTIENDFNTRLRCYFGNSWTHLQAVDWKELY 180
I+ I+ TQ+ + Q L D+LPT+E+DF L + GN+W + A +E +
Sbjct: 125 ILPISATQTAFLCRQAISIKVLRWLEDLLPTLESDFGLATMFVGNAWYQVAAGTLRECF 184

60 Query: 181 EEEYKLFITLFLSHKAEQHYCRFPKMAWLANQSPMPSIAKAKCLQHILDTSDTSAIIKAL 240
EEE +L T +L ++ + F + LW+L + ++ + Q + SD + ++ AL
Sbjct: 185 EEECQLLTAYLRQQSGRKLTLFSGMLWLSLSSHHTFLALTRQFHQFLSPQSDMADVVHAL 244

Query: 241 WQEQGNLAKTAKALFIHRNSLQYKLDKFTQSSGLNLKILDDLAYAYL 287
W E GNL +TA+ L+IHRNSLQYKLDKF Q SGL+LK LDDLA+A+L
Sbjct: 245 WSEHGNLVQTAQRLYIHRNSLQYKLDKFAQQSGLHLKQLDDLAFAHL 291

-2529-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2242

5 A DNA sequence (GBSx2362) was identified in *S.agalactiae* <SEQ ID 6931> which encodes the amino acid sequence <SEQ ID 6932>. This protein is predicted to be multiple sugar-binding transport ATP-binding protein msmk (malK). Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4392(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26938 GB:M77351 ATP-binding protein [Streptococcus mutans]

Identities = 320/377 (84%), Positives = 359/377 (94%)

20 Query: 1 MVELNLNHIYKKYPSASHYSVEDFDLDIKKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60
 MVELNLNHIYKKYP++SHYSVEDFDLDIK+KEFIVFVGPSGCGKSTTLRM+AGLEDI++G
 Sbjct: 1 MVELNLNHIYKKYPNSSHYSVEDFDLDIKNKEFIVFVGPSGCGKSTTLRMVAGLEDITKG 60

Query: 61 ELKIDGEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKFSKQEIDKRVREAA 120
 ELKIDGEVVNDK+PKDRDIAMVFQNYALYPHM+VYDNMAFGLKLR +SK+ IDKRV+EAA
 25 Sbjct: 61 ELKIDGEVVNDKAPKDRDIAMVFQNYALYPHMSVYDNMAFGLKLRHYSKEAIDKRVKEAA 120

Query: 121 ANIGLTFELERKPADLSGGQQRQVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAETIAK 180
 +GLTFELERKPADLSGGQQRQVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAETIAK
 30 Sbjct: 121 QILGLTFELERKPADLSGGQQRQVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAETIAK 180

Query: 181 IHORIGSTTIYVTHDQTEAMTLADRVIMSATKNPDGDTIGKIEQVGSPOELYNLPANK 240
 IH+RIG+TTIYVTHDQTEAMTLADRVIMS+TKN DG GTIG++EQVG+PQELYN PANK
 Sbjct: 181 IHRRIGATTIYVTHDQTEAMTLADRVIMSSTKNEDGSGTIGRVEQVGTPOELYNRPANK 240

35 Query: 241 FVAGFIGSPSMNFFKVKVENGMIISEDGLRIAIEGQEKLLSRGYKGKELIFGIRPEDI 300
 FVAGFIGSP+MNFF V +++G ++S+DGL IA+ EQG K+LES+G+K K LIFGIRPEDI
 Sbjct: 241 FVAGFIGSPAMNFFDVTKDGLVSKDGLTIAVTEGQLKMLESKGFKKNLIFGIRPEDI 300

40 Query: 301 SSNLLVQDTYPNANVEAEVLVSELLGSETMLYVKLGQTEFASRVEARDFHNPGEKVNLT 360
 SS+LLVQ+TYP+A V+AEV+VSELLGSETMLY+KLGQTEFA+RV+ARDFH PGEKV+LTF
 Sbjct: 301 SSSLLVQETYPDATVDAEVVVSELLGSETMLYKLGQTEFAARVDARDFHEPGEKVS LTF 360

Query: 361 NVAKGHFFDADETEAIR 377
 NVAKGHFFDA+TE AIR
 45 Sbjct: 361 NVAKGHFFDAETEAIR 377

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6933> which encodes the amino acid sequence <SEQ ID 6934>. Analysis of this protein sequence reveals the following:

Possible site: 48

50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4642(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 332/377 (88%), Positives = 359/377 (95%)

-2530-

- Query: 1 MVELNLNHIYKKYPSASHYSVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60
 Sbjct: 1 MVELNLNHIYKKYPNTTHYAVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60
- Query: 61 ELKIDGEVVDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKFSKQEI DKRVREAA 120
 ELKI GEVVDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRK+ K +ID+RV+EAA
 Sbjct: 61 ELKIGGEVVDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKYKDDIDRRVKEAA 120
- Query: 121 ANIGLTEFLERKPADLSGGQQRQVAMGRAIVRDAKVFLMDEPLSNLDAKL RVSMRAETIAK 180
 +GLTEFLERKPADLSGGQQRQVAMGRAIVRDAKVFLMDEPLSNLDAKL RVSMRAETIAK
 Sbjct: 121 QILGLTEFLERKPADLSGGQQRQVAMGRAIVRDAKVFLMDEPLSNLDAKL RVSMRAETIAK 180
- Query: 181 IHQRIGSTTIYVTHDQTEAMTLADRVIMSATKNPDGDTIGKIEQVGSPOELYNL PANK 240
 IH+RIGSTTIYVTHDQTEAMTLADRVIMSATKNP G+GTIGKIEQVGSPOELYNL PANK
 Sbjct: 181 IHRRIGSTTIYVTHDQTEAMTLADRVIMSATKNPQNGTIGKIEQVGSPOELYNL PANK 240
- Query: 241 FVAGFIGSPSPMNFFKVKVENGMII SEDGLRIAIPEGQEKLLSRGYKGKELIFGIRPEDI 300
 FVAGFIGSP+MNFF+V+V+G I+SEDGL IAIPEGQ K+LE+ GYKG+++ FGIRPEDI
 Sbjct: 241 FVAGFIGSPAMNFFEVEVKDGRIVSEDGLDIALPEGQAKMLEAAGYKGEKVTFGIRPEDI 300
- Query: 301 SSNLLVQDTYPNANVEAEVLVSELLGSETMLYVKLGQTEFASRVEARDFHNPGEKVN LTF 360
 SS +V DTYP+A V AEVLVSELLGSETMLYVKLGQTEFASRV+ARDFH+PGE+V+LTF
 Sbjct: 301 SSRQIVHDTYPSATVTAEVLVSELLGSETMLYVKLGQTEFASRVDARDFHSPGEQVSLTF 360
- Query: 361 NVAKGHFFDADTEQAIR 377
 NVAKGHFFD DTEQAIR
 Sbjct: 361 NVAKGHFFDRDTEQAIR 377
- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2243

A DNA sequence (GBSx2363) was identified in *S.galactiae* <SEQ ID 6935> which encodes the amino acid sequence <SEQ ID 6936>. This protein is predicted to be glucan 1,6-alpha-glucosidase (dexB) (treC).

- Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

- Final Results -----
 bacterial cytoplasm --- Certainty=0.2525(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- >GP:CAA51348 GB:X72832 glucan 1,6-alpha-glucosidase [Streptococcus
 equisimilis]
 Identities = 413/535 (77%), Positives = 476/535 (88%), Gaps = 1/535 (0%)
- Query: 1 MKKHWWHKATIIYQIYPRSFMSDGDGVGDIKGITSKLDYLEKLGITAIWLSPVYQSPMDD 60
 M+K WWHKATIIYQIYPRSF D+ G+G+GD+KGITS+LDYL+KLGITAIWLSPVYQSPMDD
 Sbjct: 1 MQKQWWHKATIIYQIYPRSFKDTSGNGIGDLKGITSQLDYLQKLGITAIWLSPVYQSPMDD 60
- Query: 61 NGYDISDYQAIADIFGDMNDMDQLLQEANQRGIKIIMDLVVNHTSDEHAWFVEARENPN 120
 NGYDISDY+AIA++FG+M+DMD LL AN+RGIKIIMDLVVNHTSDEHAWFVEARENPN
 Sbjct: 61 NGYDISDYEAIAEVFGNMDDMDLLAAANERGIKIIMDLVVNHTSDEHAWFVEARENPN 120
- Query: 121 PERDFYIWRDEPNDLTSIFSGSAWEYDKVSGQYYLHLFSKRQPDNLWENEALRHKIYDMM 180
 PERD+YIWRDEPN+L SIFSGSAWE D+ SGQYYLHLFSK+QPDNLWEN +R KIYDMM
 Sbjct: 121 PERDYIWRDEPNLMSIFSGSAWELEASGQYYLHLFSKKQPDNLWENAHVRQKIYDMM 180
- Query: 181 NFWIDKGIGGFRMDVIDLIGKIPDKGITGNGPKLHDYKEMNRASFGKHDLLTVGETWGA 240

-2531-

NFWI KGIGGFRMDVIDLIGKIPD ITGNGP+LHDYLKEMN+A+FG HD++TVGETWGA
 Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKIPDSEITGNGPRLHDYLKEMNQATFGNHDVMTVGETWGA 240

Query: 241 TPDIKQYSPNDNEELSMVFQFEHVGLQHKPDAPKWDYSDGLDVPALKDIFTKWQTQLEL 300
 TP+IA+QYS P+N+ELSMVFQFEHVGLQHKP+APKWDY++ LDVPALK IF+KWQT+L+L
 Sbjct: 241 TPEIARQYSRPNKELSMVFQFEHVGLQHKPNAPKWDYAEELDVPALKTIFSKWQTELKL 300

Query: 301 GQGWNSLFWNNHDLPRVLSIWGNDSNRKQSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360
 G+GWNSLFWNNHDLPRVLSIWGNDS R++SAKALAILLHLMRGTPYIYQGEEIGMTNYP
 Sbjct: 301 GEGWNSLFWNNHDLPRVLSIWGNDSIYREKSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360

Query: 361 FECLADVDDIESLNYAKEAMDNGVSEATILDSIRKVGGRDNARTPMQWSQEHQAGFTKG-T 419
 F+ L +VDDIESLNYAKEAM+NGV A ++ SIRKVGGRDNARTPMQWS++ AGF++
 Sbjct: 361 FKDLTEVDDIESLNYAKEAMNGVPAARVMSIRKVGGRDNARTPMQWSKDTAGFSEAQE 420

Query: 420 PWLAVNPYQEIINVEAALNDTESIFYTYQKLVALRKEHDWLVDADFKLLETADKVFAYVR 479
 WL VNPNYQEIINV AL + +SIFYTYQ+L+ALRK+ DWLV+AD+ LL TADKVFAY R
 Sbjct: 421 TWLPVNPYQEIINVADALANQDSIFYTYQQLIALRKDQDWLVEADYHLLPTADKVFAYQR 480

Query: 480 QTDKERYLIVANLSDQNQSFEPFAVKETIISNTEVQEVLSNTLKPWDAFCIEL 534
 Q +E Y+IV N+SDQ Q F A E +I+NT+V +VL + L+PWDAFC++L
 Sbjct: 481 QFGEETYVIVNVSDQEQVFAKDLAGEVVITNTDVKVLETKHLQPWDAFCKVL 535

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6937> which encodes the amino acid
 sequence <SEQ ID 6938>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2793 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 418/535 (78%), Positives = 474/535 (88%), Gaps = 1/535 (0%)

Query: 1 MKKHWWHKATIYQIYPRSFMDSDGVDGDIKGITSKLDYLEKLGITAIWLSPVYQSPMDD 60
 M HWWHKATIYQIYPRSF D+ G+G+GD+KGITS+LDYL+KLGITAIWLSPVYQSPMDD
 Sbjct: 1 MNNHWWHKATIYQIYPRSFKDTSGNGIGDLKGITSQLDYLQKLGITAIWLSPVYQSPMDD 60

Query: 61 NGYDISDYQAIADIFGDMNDMDQLLQEANQRGIKIIMDLVVNHTSDEHAWFVEARENPN 120
 NGYDISDY+ATAD+FGDM DMD+LL AN+RGIKIIMDLVVNHTSDEHAWFVEARENPN
 Sbjct: 61 NGYDISDYEAADVFGDMDMDDELLAAANERGKIIMDLVVNHTSDEHAWFVEARENPN 120

Query: 121 PERDFYIWRDEPNLTSIFSGSAWEYDKVSGQYYLHLFSKRQPDNLWENALRHKIYDMM 180
 PERD+YIWRDEPN+L SIFSGSAWE D+ SGQYYLHLFSK+QPDNLWEN LR KIYDMM
 Sbjct: 121 PERDYIWRDEPNLMSIFSGSAWELDEASGQYYLHLFSKKQPDNLWENAQLRQKIYDMM 180

Query: 181 NFWIDKGIGGFRMDVIDLIGKIPDKGITGNGPKLHDYLKEMNRASFGKHDLLTVGETWGA 240
 NFWI KGIGGFRMDVIDLIGK+PD ITGNGP+LHDYLKEMN+A+FG HD++TVGETWGA
 Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKVPDLEITGNGPRLHDYLKEMNQATFGNHDVMTVGETWGA 240

Query: 241 TPDIKQYSPNDNEELSMVFQFEHVGLQHKPDAPKWDYSDGLDVPALKDIFTKWQTQLEL 300
 TP+IA+QYS P+N+ELSMVFQFEHVGLQHKPDAPKWDY+ LDVPALK IF+KWQT+L+L
 Sbjct: 241 TPEIARQYSRPNKELSMVFQFEHVGLQHKPDAPKWDYAKELDVPALKAFSKWQTELKL 300

Query: 301 GQGWNSLFWNNHDLPRVLSIWGNDSNRKQSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360
 G+GWNSLFWNNHDLPRVLSIWGNDS R++SAKALAILLHLMRGTPYIYQGEEIGMTNYP
 Sbjct: 301 GEGWNSLFWNNHDLPRVLSIWGNDSIYREKSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360

Query: 361 FECLADVDDIESLNYAKEAMDNGVSEATILDSIRKVGGRDNARTPMQWSQEHQAGFTKG-T 419
 F+ L +V+DIESLNYAKEAM NGVS A ++DSIRKVGGRDNARTPMQWS++ AGF++
 Sbjct: 361 FKDLTEVNDIESLNYAKEAMNGVSAARVMSIRKVGGRDNARTPMQWSKDTAGFSEAKE 420

Query: 420 PWLAVNPYQEIINVEAALNDTESIFYTYQKLVALRKEHDWLVDADFKLLETADKVFAYVR 479

-2532-

WL VNPNYQ+INV AL D +SIFYTYQKL+ALRKE DWLV+AD+ LL TADKVFAY R
 Sbjct: 421 TWLPVNPNYQDINVADALADPDSIFYTYQKLIALRKEQDWLVEADYHLLPTADKVFAYQR 480

Query: 480 QTDKERYLIVANLSDQNQSFEFPPEAVKETIISNTEVQEVLSNLTLPWDAFCIEL 534
 Q +E Y+IV N+SD+ Q F A + II+NT+V VL + L+PWDAF++L
 Sbjct: 481 QLGEETYVIVVNVSDDEEQVFATDLAGAQVLIANTDVTVLTKHLQPWDAFCLKL 535

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 2244

A DNA sequence (GBSx2364) was identified in *S. agalactiae* <SEQ ID 6939> which encodes the amino acid sequence <SEQ ID 6940>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB49738 GB:U21942 UDP-galactose 4-epimerase [Streptococcus mutans]
 Identities = 267/331 (80%), Positives = 306/331 (91%)

Query: 1 MAVLILGAGYIGSHMVDQLITQGKEKVIIVVDNLVTGHRQAVHSDAIFYEGDLSDKTFMR 60
 MA+L+LGGAGYIGSHMVD+LI +G+E+V+VVD+LVTGHR AVH A FY+GDL+D+ FM
 Sbjct: 1 MAILVLGGAGYIGSHMVDRLIEKGEEVVVVDLSLVTGHRAAVHPAKFYQGDLADREFMS 60

Query: 61 QVRENPDVDAVIHFAAFSLVAESMENPLKYFDNNTAGMIKLEVMNECDIKNIVFSSTA 120
 VVRENPDVDAVIHFAA+SLVAESM+ PLKYFDNNTAGMIKLEVM+E +K IVFSSTA
 Sbjct: 61 MVFRENPDVDAVIHFAAYSLVAESMKKPLKYFDNNTAGMIKLEVMSEFGVKYIVFSSTA 120

Query: 121 ATYGIFEQVPILETAPQNPINPYGESKLMMETIMKWADQAYGIKFVALRYFNVAGDKPDG 180
 ATYGIP ++PI ET PQ PINPYGESKLMMETIMKW+D+AYGIKFV +RYFNVAG KPDG
 Sbjct: 121 ATYGIPNEIPIKETTPQRPINPYGESKLMMETIMKWSDRAYGIKFVPVRYFNVAGAKPDG 180

Query: 181 SIGEDHKKPETHLLPIILQVAQGVDRKIMIFGDDYNTPDGTNVRDYYHPPDLADAHILAVD 240
 SIGEDH PETHLLPIILQVAQGV+KIMIFGDDYNTPDGTNVRDYYHPPDLAD H+LA++
 Sbjct: 181 SIGEDHSPETHLLPIILQVAQGVREKIMIFGDDYNTPDGTNVRDYYHPPDLADRHLLALN 240

Query: 241 YLRQGNESNVFNLSSTGFSNLQMLEAARRITGKEIPAQKAARRGDPDTLIASSEKARQ 300
 YLRQGN S FNLGSSTGFSNLQ+LEAAR++TG++IPA+KAARR GDPDTLIASSEKAR+
 Sbjct: 241 YLRQGNPSTAFNLSSTGFSNLQILEAARKVTGQKIPAQKAARRSGDPDTLIASSEKARE 300

Query: 301 ILGWEPKFDNIDKIISAWAWHSSHPNGYED 331
 ++GW+P+FD+I+KII+SAWAWHSSHP GY+D
 Sbjct: 301 VVGWKEQFDDIEKIIASAWAWHSSHPKGYDD 331

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2245

A DNA sequence (GBSx2366) was identified in *S. agalactiae* <SEQ ID 6941> which encodes the amino acid sequence <SEQ ID 6942>. This protein is predicted to be two-component response regulator. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

-2533-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3945(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BAB06470 GB:AP001516 two-component response regulator [Bacillus halodurans]
 Identities = 71/223 (31%), Positives = 139/223 (61%), Gaps = 7/223 (3%)

Query: 3 VLIIEDDPMVEFIHRNYLEKLNLYFQNIYSTASQTQAIAYLNDIKIQLVLLDIHIKEGNGL 62
 VL+IEDDPMV+ ++R ++EKL+ F + +TA+ + + +++ L+LLDI + + +GL
 Sbjct: 9 VLIIEDDPMVQEVNRMFVEKLSGFTIVGTATGEEGMVKPRELQPDLLILDIFMPKQDGL 68

15 Query: 63 ELLKLLRNQHONTEVIVISAANEHTVKEAFHLGIVDYLIKPFTEFFERFESSIEKFLNHYH 122
 +K +R Q+ + ++I ++AAN+ T+K G++DYL+KPFTFER +++++ + +
 Sbjct: 69 SFIKQIREQYIDVDIIAVTAANDTKTIKTLRLRYGVM DYLVKPFTEFLKAALTQYEEMPR 128

20 Query: 123 TFEAD-KIYQDNIDHFQKIDSGWLEGEVKLDE--KGLSEITYQHILDAIQELEQPFTIQE 179
 + + ++ QD++D K + + +D+ KGL T Q +++ ++EL++P + +E
 Sbjct: 129 KMQKEAELSQDSLDEMIK----QKQAQANMDDLPGKGLHAHTLQQVIERLEELDEPKSABE 184

25 Query: 180 LAKCSQFSHVSVRKYIAYMEEKGLLTSQQIYTKVGRPYKVYKL 222
 + + + V+VR+Y+ Y+E G + Y +GRP + YKL
 Sbjct: 185 IGRDVGLARVTVRRYLYNLESVGQVEMDLTYGSIGRPIQTYKL 227

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6943> which encodes the amino acid sequence <SEQ ID 6944>. Analysis of this protein sequence reveals the following:

30 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.4053(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 123/220 (55%), Positives = 156/220 (70%)

40 Query: 1 MDVLIIEDDPMVEFIHRNYLEKLNLYFQNIYSTASQTQAIAYLNDIKIQLVLLDIHIKEGN 60
 M+VLIIEDDPMV+FIHRNYLEKLN F I S+ S + L D I L+LLDIHI +GN
 Sbjct: 1 MNVLIIEDDPMVDFIHRNYLEKLNLFDRITSSDSMKAVQSILTDYAIIDLILLDIHITDGN 60

45 Query: 61 GLELLKLLRNQHONTEVIVISAANEHTVKEAFHLGIVDYLIKPFTEFFERFESSIEKFLNH 120
 G++ L+ R QH EVI+ISAAN+ + +++ PHLGI+DYLIKPFTEFFERF+ SI++F+ H
 Sbjct: 61 GIQFLEKWRTOHIPCEVIIISAANDGNIIRDGFHLGIIDYLIKPFTEFFERFQESIQQFVTH 120

50 Query: 121 YHTFEADKIYQDNIDHFQKIDSGWLEGEVKLDEKGLSEITYQHILDAIQELEQPFTIQEL 180
 ++ Q ID + + S +L EKGLSE T+Q I++ I+ +QPFTIQEL
 Sbjct: 121 REHLANQQLEQAQIDQLKCLTSKKDTKNKQLEKGLSESTFQWIMENIKVFDQPFTIQEL 180

55 Query: 181 AKCSQFSHVSVRKYIAYMEEKGLLTSQQIYTKVGRPYKVY 220
 A SHVSVRKYIAY+EE L SQI+TKVGRPY+VY
 Sbjct: 181 ASACHLSHVSVRKYIAYLEENKQLNSQQIFTKVGRPYRVY 220

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2246

60 A DNA sequence (GBSx2367) was identified in *S.agalactiae* <SEQ ID 6945> which encodes the amino acid sequence <SEQ ID 6946>. Analysis of this protein sequence reveals the following:

-2534-

Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.76	Transmembrane	12 - 28 (6 - 34)
INTEGRAL	Likelihood = -7.43	Transmembrane	178 - 194 (173 - 197)

----- Final Results -----

bacterial membrane	---	Certainty=0.4503(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9003> which encodes amino acid sequence <SEQ ID 9004> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 3

SRCFLG: 0

McG: Length of UR: 27

Peak Value of UR: 2.99

Net Charge of CR: 3

McG: Discrim Score: 12.92

GvH: Signal Score (-7.5): -2.57

Possible site: 19

>>> Seems to have an uncleavable N-term signal seq

Amino Acid Composition: calculated from 1

ALOM program count: 2 value: -8.76 threshold: 0.0

INTEGRAL Likelihood = -8.76 Transmembrane 10 - 26 (4 - 32)

INTEGRAL Likelihood = -7.43 Transmembrane 176 - 192 (171 - 195)

PERIPHERAL Likelihood = 3.18 149

modified ALOM score: 2.25

icml HYPID: 7 CFP: 0.450

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane	---	Certainty=0.4503(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15141 GB:Z99120 similar to two-component sensor histidine
kinase [YuFM] [Bacillus subtilis]

Identities = 132/461 (28%), Positives = 245/461 (52%), Gaps = 7/461 (1%)

Query: 3 MKKLSLWAFSLILVTMTICIFSIYYVTIHQSRYMVRVQEEKILKNTGYALSRNPQVI 62
MKK L L L++ + + + I ++ Q+ + +R QE+ T ++ P

Sbjct: 1 MKKTLKLQTRLTIFVCIVVLIALLLITFFTVGAQTTRKIRDOEKATLQTAEMVAEAPMTA 60

Query: 63 QTLKDNHYDQSLQKQMLFLSKKSNLDYIVLINLKGIRFTHPDSTKIGKPFQGGDEQAVFK 122
L+ + LQ + K + +++V++++ GIR THPD +KIGK F+GGDE V K

Sbjct: 61 AALESQKQKELQSYTKRVQKITGTEFVVMDMNGIRKTHPDPSKIGKKFRGGDESEVLK 120

Query: 123 GKAIMSTAEGSLGKSLRYLIPVY-DHQKQVGAIAGVLKLTTLGDLSQSSIKEFSKPLLIS 181
G +STA G+LGKS R +PVY ++ KQVGA+AVG+ + + ++ S++ + +S

Sbjct: 121 GHVHISTASGTLGKSQRAFPVYAENGKQVGAVAVGITVNEIDEVISHSLRPLYFIICVS 180

Query: 182 ILISLVVTSIISYGLKKQLHNLHPSDIFQHLERNATLDQIAAVFVIDQRHIIKRNNPA 241
I + ++ I++ +K ++ L P +I LEER+A L+ + + +D+ IK N

Sbjct: 181 IFVGIVGAVIVARTVKNIMYGLEPYEIAITLLEERSAMLESTKEGILAVDEHGKIKLANAE 240

Query: 242 ASLLFKKEGQDRLFSKGLLESLLIP--QLKQDHFSKK--TEQVLHFQGGDYLLSISPI TVK 297
A L F K G + ++ ++P +LK+ +KK ++ + G + + + PI +K

Sbjct: 241 AKRLFVKMGININPIDQDQDDILPKSRLKKVIETKKPLQDRDVRINGLELVFNEVP IQLK 300

Query: 298 TQNRGYVVFRLRNVTETLFTLDQLAHTTAYASALQAQTHQFMNQLHVIYGLADIEYYDELK 357
Q G + R+ TE +QL+ YA+AL+AQ+H+FMN+LHVI GL ++ YD+L

Sbjct: 301 QGTVGAIATFRDKTEVKHLAEQLSGVKMYANALRAQSHEFMNKLHVILGLVQLKEYDDLK 360

-2535-

Query: 358 IYLKELLEPQNEFLARLSMLVREPRLASFIIGEREKFAEKHINLSTEILVEIPTKSTVED 417
 Y+K++ Q + + V+ LA F++G++ E+ NL E IP +
 Sbjct: 361 DYIKDIAIQKSETSETINDVKSSVLGFLGKQSFIREQGANLDIECNGVIPNAADPSV 420

5 Query: 418 VNNYL-LLHRYINTKILTLN-STTLVSLRLNYQNNLIETD 456
 ++ + ++ IN + + + +++ + + N++++ +
 Sbjct: 421 IHELITTIIGNLINNGLDAVADMPKKQITMSMRFHNSILDIE 461

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6947> which encodes the amino acid
 10 sequence <SEQ ID 6948>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -10.03 Transmembrane 174 - 190 (170 - 195)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 236/488 (48%), Positives = 337/488 (68%), Gaps = 3/488 (0%)

Query: 3 MKKKLSLWAFSLILVTMTICIFSIFFYVTHQSYRMVRVQEEKILKNTGYALSRNPQVI 62
 MKK L LWA LSLILV+M + S+FY + +H +++ ++ QE +L +TG L+ + +
 25 Sbjct: 1 MKKPLRLWASLSLILVSMIVVTTSLFYGIMLHDTHQSIKNQETHLLTSTGKMLASHQAIAK 60

Query: 63 QTLKDNHYDQSLQKQMLFSLKSNLDYIVLINLKGIRFTHPDSTKIGKPFQGGDEQAVFK 122
 + L +N + ++ NLDY+V++N+KGIR THP+ ICKPFQGGDE+AV
 30 Sbjct: 61 ELLLNQPNNAKTAYTNSIASIYNLDYVVMNMKGIRLTHPNKNIGKPFQGGDEEAVLA 120

Query: 123 GKAIMSTAEGSLGKSLRYLIPVYDHQKQVGAIAVGLKLTTLGDLSQSSIKFSPKLLISI 182
 GK ++STA+G+LGKSLRYL+PV+D KQ+GAIAVG+KLTTL D++ +S + ++ LL+ +
 35 Sbjct: 121 GKKVISTAKGTGKSLRYLVPVFDGDKQIGAIAGIKLTTLNDVALTSKRNYTLSELLCL 180

Query: 183 LISLVVTSIISYGLKKQLHNLHPSDIFQHLEERNATLDQIQAAVVIDQRHIIKRNNPAA 242
 LISL+VTS IS+ LK+QLH L PS+I+Q EERNA LDQI+AAVFD+ I++ N A
 40 Sbjct: 181 LISLVTSFISFRLKQLHQLPSEIYQLFEERNAMLDQIEAAVFDKAGILQLCNQAG 240

Query: 243 SLLFKKEGQDRLFSGLLESILPQLKQDHFSKKTQVLFHFQGDYLLSISPITVKTQNRG 302
 L ++ Q +G L P + + EQ+ + +DYLL+ISPI VK +RG
 45 Sbjct: 241 QKLIARKQLGKPTGNSFNLYFPDFPKLSLQEGHEQLFRYGEEDYLLAISPICVKNDRHG 300

Query: 303 YVFLFRNVTTETLFTLDQLAHTTAYASALQAQTHQFMNQLHVIYGLADIEYDELKIYLKE 362
 +++F+R + + TLDQLA+TTAYASALQAQTH+FMNQLHVIYGL DI YYD+LKIYL
 50 Sbjct: 301 HIIFMREAVKAIDTLDQLAYTTAYASALQAQTHQFMNQLHVIYGLVDIAYDQLKIYLD 360

Query: 363 LLEPQNEFLARLSMLVREPRLASFIIGEREKFAEKHINLSTEILVEIPTKSTVEDVNNYL 422
 +LEP+NE L LS+LV+EP LASF+IGE+EK+ E +++L ++L EIP +T +NN I
 55 Sbjct: 361 ILEPENEILTSLSVLKEPLLASFLIGEKEKYQELNVHLKIDVLSEIPHSATKNQLNNGL 420

Query: 423 LLHRYINTKILTLNLTTLVSLRLNYQNNLIETDYQWENEKWL-LNDYHQYFNDAYFQQL 481
 +++R+I+T +LT L +LV + QN+LI + + W+ L F+ YFQQL
 60 Sbjct: 421 MIYRFIHTNLLTTLRPKSLVLSIQHDQNHLI--SHYTLTDNWIDLERVQPIFDLPYFQQL 478

Query: 482 LVDSRATY 489
 L D+ + +
 65 Sbjct: 479 LTDINSQF 486

SEQ ID 9004 (GBS130d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total
 60 cell extract is shown in Figure 123 (lane 8-10; MW 63kDa) and in Figure 184 (lane 4; MW 63kDa). It was
 also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure
 123 (lane 11; MW 38kDa) and in Figure 181 (lane 7; MW 38kDa).

-2536-

GBS130d-GST was purified as shown in Figure 237, lane 11. GBS130d-His was purified as shown in Figure 233, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 2247

A DNA sequence (GBSx2368) was identified in *S.agalactiae* <SEQ ID 6949> which encodes the amino acid sequence <SEQ ID 6950>. Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence
10  INTEGRAL    Likelihood = -11.52    Transmembrane  364 - 380 ( 353 - 386)
    INTEGRAL    Likelihood = -9.66     Transmembrane   33 - 49 ( 26 - 57)
    INTEGRAL    Likelihood = -7.80     Transmembrane   87 - 103 ( 82 - 105)
    INTEGRAL    Likelihood = -6.85     Transmembrane  153 - 169 ( 144 - 174)
    INTEGRAL    Likelihood = -4.41     Transmembrane  301 - 317 ( 300 - 318)
15  INTEGRAL    Likelihood = -2.81     Transmembrane  216 - 232 ( 212 - 235)
    INTEGRAL    Likelihood = -2.39     Transmembrane  120 - 136 ( 120 - 136)
    INTEGRAL    Likelihood = -1.65     Transmembrane   57 - 73 ( 56 - 73)
    INTEGRAL    Likelihood = -1.17     Transmembrane  428 - 444 ( 428 - 444)
20  INTEGRAL    Likelihood = -0.32     Transmembrane  276 - 292 ( 276 - 292)

----- Final Results -----
        bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB18291 GB:U35658 L-malate permease [Streptococcus bovis]
Identities = 329/428 (76%), Positives = 375/428 (86%)

30  Query: 18  DLKAKLFHIKIGSVPLPVYVCLALLILLAGFLQKLPVNMLGGFAVILTMGWFLGTIGASI 77
      D + KL  +IGSV LPVY+ A +IL+ L++LPVNMLGGFAVILTMGW LGTIG +I
      Sbjct: 14  DWRNKLTKTRIGSVTLPVYLVLTASIIILVTALLEQLPVNMLGGFAVILTMGWLLGTIGGNI 73

      Query: 78  PGFKNFGGPAILSLVPSILVFFNLINKNVLESTNMLMKQANFLYFYIACLVSGSILGMN 137
      P K+FGGPAILSLVPSI+VFFNL+N+NVL+ST++LMKQANFLYFYIACLV GSILGMN
35  Sbjct: 74  PILKHFGGPAILSLVPSIMVFFNLLNQNVLDSTDILMKQANFLYFYIACLVCGSILGMN 133

      Query: 138 RKMLIQGLLRMIFPMLLGMVCMVMVGTFVGVILGLEWRHTLFYIVTPVLGGIGEGILPL 197
      RK+L+QGL+RMI PM LGM+ AM VGT VG +LGL W+H+LFYIVTPVLGGIGEGILPL
40  Sbjct: 134 RKILVQGLMRMIVPMALGMILAMGVGTLVGTLLGLGWKHSLFYIVTPVLGGIGEGILPL 193

      Query: 198 SLGYSSITGVASEQLVAQLIPATIIGNFFAILCTALLNRLGEKKPHLSGQQLVRLNKGE 257
      SLGYS+ITG+ SEQLV QLIPATIIGNFFAI+C+ LL+RLGEK+P LSGQQL+++ +
45  Sbjct: 194 SLGYSAITGLPSEQLVGQLIPATIIGNFFAIMCSGLLSRLGEKRPESLQQLIKITNSD 253

      Query: 258 DMSDIIADHSGPIDVKMKGGGVLTA CSLFIFGHLLQQLTGFPGPVLMIVAAAILKYINVI 317
      D+SD + + PIDVK MG GVL AC+LFI G LLQ LTGFPGPVLMIV AA LKY+NV+
      Sbjct: 254 DLSDALEEDKAPIDVKLMGAGVLIACTLFITGGLLQHLTGFPGPVLMIVAAFLKYNLV 313

50  Query: 318 PRETONGAKQLYKFISGNFTFPLMAGLGLLYIPLKDVVATLSIQYFIVVISVVFVVISVG 377
      P+ETQ G+KQLYKFISGNFTFPLM GLG+LYIPLKDVV LS QYF+VVISVVFVI+ G
      Sbjct: 314 PKETQRGSKQLYKFISGNFTFPLMVGLGMLYIPLKDVVGMLSWQYFVVVISVVFVIATG 373

      Query: 378 FFVSRFLNMNPVEAGIISACQSGMGGTGDVAILSTADRMNLMPPAQVATRLGGAITVITM 437
      FFVSRF+NMNPVEA I+SACQSGMGGTGDVAILSTA+RM LMPFAQVATRLGGAITVITM
55  Sbjct: 374 FFVSRFNMNPVEAAIVSACQSGMGGTGDVAILSTANRMTLMPFAQVATRLGGAITVITM 433

      Query: 438 TAILRMLF 445
      TAI RMLF
60  Sbjct: 434 TAIFRMLF 441

```

-2537-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6951> which encodes the amino acid sequence <SEQ ID 6952>. Analysis of this protein sequence reveals the following:

```

Possible site: 48
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.89    Transmembrane    361 - 377 ( 350 - 383)
    INTEGRAL    Likelihood = -7.43     Transmembrane     84 - 100 ( 79 - 102)
    INTEGRAL    Likelihood = -6.16     Transmembrane    150 - 166 ( 137 - 171)
    INTEGRAL    Likelihood = -4.88     Transmembrane     30 - 46 ( 24 - 48)
10  INTEGRAL    Likelihood = -4.35     Transmembrane    299 - 315 ( 297 - 316)
    INTEGRAL    Likelihood = -4.14     Transmembrane    117 - 133 ( 115 - 134)
    INTEGRAL    Likelihood = -3.19     Transmembrane     54 - 70 ( 51 - 75)
    INTEGRAL    Likelihood = -2.92     Transmembrane    425 - 441 ( 425 - 442)
    INTEGRAL    Likelihood = -2.81     Transmembrane    213 - 229 ( 209 - 232)
15  INTEGRAL    Likelihood = -2.44     Transmembrane    273 - 289 ( 271 - 290)

----- Final Results -----
    bacterial membrane --- Certainty=0.5755(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAB18291 GB:U35658 L-malate permease [Streptococcus bovis]
Identities = 344/443 (77%), Positives = 394/443 (88%), Gaps = 6/443 (1%)
25  Query: 4   ISKKMPQKDLSEHSAWQNR---RIGSVPLPVYLVLATLILVTGWLQQLPVNMLGGFAV 59
    + KK+P    +E    W+N+    RIGSV LPVYL V A++ILVT  L+QLPVNMLGGFAV
    Sbjct: 1   MEKKLPATAANETD--WRNKLTKTRIGSVTLFPVYLVTASIIILVTALLEQLPVNMLGGFAV 58

30  Query: 60  ILTLGWLLGTIGATIPGLKHFGGPAILSLVPSILVFFNLLNPNVLEATNVLMKQANFLY 119
    ILT+GWL LGTIG IP LKHFGGPAILSLVPSI+VFFNLLN NVL++T++LMKQANFLY
    Sbjct: 59  ILTMGWLLGTIGGNIPILKHFGGPAILSLVPSIMVFFNLLNQNVLDSTDILMKQANFLY 118

35  Query: 120 FYIACLVCGSILGMNRKILIQGLFRMIIPMLLGMVCAMGVGTLVGVLGLDWQHTLFYVV 179
    FYIACLVCGSILGMNRKIL+QGL RMI+PM LCM+ AMGVGTLVG +LGL W+H+LFY+V
    Sbjct: 119 FYIACLVCGSILGMNRKILVQGLMRMIVPMALGMILAMGVGTLVGTLGLGWKHSIFYIV 178

40  Query: 180 TPVLAGGIGEGILPLSLGYSAITGVGSEQLVQLIPATIIIGNFFAILCTALLNRFGEKHP 239
    TPVLAGGIGEGILPLSLGYSAITG+ SEQLV QLIPATIIIGNFFAI+C+ LL+R GEK P
    Sbjct: 179 TPVLAGGIGEGILPLSLGYSAITGLPSEQLVGLIPATIIIGNFFAIMCSGLLSRLGEKRP 238

45  Query: 240 SYSGQGQLVKIGHSEDMSDALKDNSGALDVKLMGAGVLTACSLFIAGGLLQHLTDFPGPV 299
    SQQGQL+KI +S+D+SDAL+++ +DVKLMGAGVL AC+LFI GGLLQHLT FPGPV
    Sbjct: 239 ELSGQGQLIKITNSDDLSDALEEDKAPIDVKLMGAGVLIACLTFTIGGLLQHLTGFPGPV 298

50  Query: 300 LMIILAAFLKYLNVIPQETQNGAKQLYKFISNFTFPLMAGLGLLYIPLKEVVATLSWQY 359
    LMI++AAFLKYLNV+P+ETQ G+KQLYKFIS NFTFPLM GLG+LYIPLK+VV LSWQY
    Sbjct: 299 LMIVVAAFLKYLNVVPKETQRGSKQLYKFISGNFTFPLMVGLGMLYIPLKDVVGMLSWQY 358

55  Query: 420 QVATRLGGAITVITMTAILRIIF 442
    QVATRLGGAITVITMTAI R++F
    Sbjct: 419 QVATRLGGAITVITMTAIFRMLF 441

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 356/419 (84%), Positives = 385/419 (90%)
60  Query: 27  KIGSVPLPVYVCLALLILLAGFLQKLPVNMLGGFAVILTMGWFLGTIGASIPGFKNFGGP 86
    +IGSVPLPVY+ LA LIL+ G+LQ+LPVNMLGGFAVILT+GW LGTIGA+IPG K+FGGP
    Sbjct: 24  RIGSVPLPVYLVLATLILVTGWLQQLPVNMLGGFAVILTLGWLLGTIGATIPGLKHFGGP 83

```

-2538-

Query: 87 AILSLLVPSILVFFNLLINKNVLESTNMLMKQANFLYFYIACLVSGSILGMNRKMLIQGLL 146
 AILSLLVPSILVFFNLL+N NVLE+TN+LMKQANFLYFYIACLV GSILGMNRK+LIQGL
 Sbjct: 84 AILSLLVPSILVFFNLLNPVLEATNVLMMKQANFLYFYIACLVCGSILGMNRKILIQGLF 143

5 Query: 147 RMIFPMLLGMVCAMMVGTFTVGVLGLLEWRHTLFYIVTPVLGGIGEGILPLSLGYSSITG 206
 RMI PMLLGMVCAM VGT VGVILGL+W+HTLFY+VTPVLGGIGEGILPLSLGYS+ITG
 Sbjct: 144 RMIIPMLLGMVCAMGVGTFLVGVLGLDWQHTLFYVVTVPVLGGIGEGILPLSLGYSAITG 203

10 Query: 207 VASEQLVAQLIPATTIIGNFFAILCTALLNRLGEKKPHLSGQQLVRLNKGEDMSDIIADH 266
 V SEQLVAQLIPATTIIGNFFAILCTALLNR GEK P SGQQLV++ EDMSD + D+
 Sbjct: 204 VGSEQLVAQLIPATTIIGNFFAILCTALLNRFGEKHPSYSGQQLVKIGHSEDMSDALKDN 263

15 Query: 267 SGPIDVKKMGGVLTACSLFIFGHLLQQLTGFPFVLMIVAAAILKYINVIPRETQNGAK 326
 SG +DVK MG GVLTAACSLFI G LLQ LT FPGFVLM+ AA LKY+NVIP+ETQNGAK
 Sbjct: 264 SGALDVKLMGAGVLTACSLFIAGLLQHLTDFFPGFVLMILAAFLKYLNVIPQETQNGAK 323

20 Query: 327 QLYKFISGNFTFPLMAGLGLLYIPLKDVVATLSIQYFIVVISVVFVISVGGFFVSRFLNM 386
 QLYKFIS NTFPLMAGLGLLYIPLK+VVATLS QYFIVVISV TV+SVGGFFVSRFLNM
 Sbjct: 324 QLYKFISNFTFPLMAGLGLLYIPLKEVVATLSWQYFIVVISVVLTVVSVGGFFVSRFLNM 383

Query: 387 NPVEAGIISACQSGMGGTGDVAILSTADRMNLMPPAQVATRLGGAITVITMTAILRMLF 445
 +PVEA IISACQSGMGGTGDVAILSTADRMNLMPPAQVATRLGGAITVITMTAILR++F
 Sbjct: 384 SPVEAAIISACQSGMGGTGDVAILSTADRMNLMPPAQVATRLGGAITVITMTAILRIIF 442

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2248

A DNA sequence (GBSx2369) was identified in *S.agalactiae* <SEQ ID 6953> which encodes the amino acid sequence <SEQ ID 6954>. This protein is predicted to be malic enzyme (mae). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.28 Transmembrane 164 - 180 (164 - 181)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB07709 GB:U35659 malic enzyme [Streptococcus bovis]
 Identities = 285/386 (73%), Positives = 332/386 (85%), Gaps = 1/386 (0%)

45 Query: 2 SENLQQLAINQARENGGKLEVISKVKVEDKRDLSIAYTFGVASVSSAIAEDVELAYELTT 61
 ++++ +LAI QA++ GKKLEV KV +E K DL IAYTFGVA+VSSAI E E AYELTT
 Sbjct: 3 TKDVKELAI EQAKKFGGKLEVCPKVPIETKADLGIAITFGVA VSSAIYEKKERAYELTT 62

Query: 62 KKNTVAVVSDGS AVLGLGDIGPEAAMPVMEGKAALFKRFANVDAVPIVLKNTDTEEIIISI 121
 KKNTVAV+SDGS AVLGLG+IGPEAAMPVMEGKAALFKRFA VD++P+VL T DTEEII
 50 Sbjct: 63 KKNTVAVISDGS AVLGLGNIGPEAAMPVMEGKAALFKRFAGVDSIPLVLDTQDTEEIIQT 122

Query: 122 VKAISPTFGGINLEDISAPRCFEIEQRLIEECDIPVFHDDQHGTAVVLAALFNSLKLK 181
 VK ++PTFGGINLEDISAPRCFEIEQRLI+E DIPVFHDDQHGTAVVLAAL+NSLKL+
 Sbjct: 123 VKFLAPTFGGINLEDISAPRCFEIEQRLIDELDIPVFHDDQHGTAVVLAALYNSLKLIN 182

55 Query: 182 KDIEDIRVVVNGGGSAGLSITRKLLSAGAKHVTVDVDFGIINDKDRESLAPHHKAIKLT 241
 K IEDI VV+NGGGSAGLSITRK L+AG KH+ +VDR GI+++ D +L PHH IAKLT
 Sbjct: 183 KKIEDIHVVINGGGSAGLSITRKFLAAGVKHIIIVDRTGILSETD-TALPHHAEIAKLT 241

60 Query: 242 NREFQSGSLEDALENADVFIGVSAPEALHAEWISKMA DKPIVFAMANPIPETYPDQALKA 301
 NRE ++G L ALE ADVF+GVSA P L EWI +M ++P++FAMANP+PEI+PD+AL A

-2539-

Sbjct: 242 NREHRTGDLATALEGADVFGVSAPGVLKPEWIIQQMNEQPVIFAMANPVPEIFPDEALAA 301

Query: 302 GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIASLIPEEELST 361
GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAK IT+EMQIAAA+GIA LIP+ EL+

5 Sbjct: 302 GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIAKLIPDNELTP 361

Query: 362 THIIIPNAFQNDVADVVAKSVSNAVQK 387
T+IIP+ FQ VA VVA+SV NAV++

10 Sbjct: 362 TNIIIPDPFQEGVAKVVAESVRNAVKE 387

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6955> which encodes the amino acid sequence <SEQ ID 6956>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

15 INTEGRAL Likelihood = -2.44 Transmembrane 164 - 180 (164 - 181)
INTEGRAL Likelihood = -1.75 Transmembrane 94 - 110 (94 - 110)

----- Final Results -----

20 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB07709 GB:U35659 malic enzyme [Streptococcus bovis]

25 Identities = 289/379 (76%), Positives = 334/379 (87%), Gaps = 1/379 (0%)

Query: 7 QLALEQAKTFGGKLEVPKVDIKTKHDLIAIYTPGVASVSSAIAKDKTLAYDLTTKKNTV 66
+LA+EQAK FGGKLEV PKV I+TK DL IAYTPGVA+VSSAI + K AY+LTTKKNTV

30 Sbjct: 8 ELAIEQAKKFGGKLEVPKVIETKADLGIAIYTPGVAAVSSAIYEKKERAYELTTKKNTV 67

Query: 67 AVISDGTAVLGLGDIGPEAAMPVMEGKAALFKAFAGVDAIPIVLDTKDTTEEIIISIVKALA 126
AVISDG+AVLGLG+IGPEAAMPVMEGKAALFK FAGVD+IP+VLDT+DTEEII VK LA

Sbjct: 68 AVISDGSAVLGLGNIGPEAAMPVMEGKAALFKRFAGVDSIPLVLDTQDTEEIIQTVKFLA 127

35 Query: 127 PTFGGINLEDISAPRCFEIEQRLIKECHIPVFHDDQHGTAVVLAALFNSLKLKSLDE 186
PTFGGINLEDISAPRCFEIEQRLI E IPVHDDQHGTAVVLAAL+NSLKL+ K +++

Sbjct: 128 PTFGGINLEDISAPRCFEIEQRLIDELDIPVFHDDQHGTAVVLAALYNSLKLINKKIED 187

40 Query: 187 VSIVVNGGGSAGLSITRKLAAAGATKVTVDKFGIINEQEAAQLAPHHLDIKVTNREFK 246
+ +V+NGGGSAGLSITRK LAAG + +VD+ GI++E + A L PHH +IAK+TNRE +

Sbjct: 188 IHVVINGGGSAGLSITRKFLAAGVKHIIIVDRTGILSETDTA-LPPHAEIAKL/TNREHR 246

Query: 247 SGTLEDALEGADIFIGVSAPGVLKAEWISKMAARPVIFAMANPIPEIYPDEALEAGAYIV 306
+G L ALEGAD+F+GVSAPGVLK EWI +M +PVIFAMANP+PEI+PDEAL AGAYIV

45 Sbjct: 247 TGDALATALEGADVFGVSAPGVLKPEWIIQQMNEQPVIFAMANPVPEIFPDEALAAGAYIV 306

Query: 307 GTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIASLVPDDALSTTNIIIP 366
GTGRSDFPNQINNVLAFPGIFRGALDARAK IT+EMQIAAARGIA L+PD+ L+ TNIIIP

50 Sbjct: 307 GTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIAKLIPDNELTPTNIIIP 366

Query: 367 DAFKEGVAEIVAKSVRSV 385
D F+EGVA++VA+SVR+ V

Sbjct: 367 DPFQEGVAKVVAESVRNAV 385

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 306/387 (79%), Positives = 349/387 (90%)

Query: 1 MSENLGQLAINQARENGGKLEVISKVKVEDKRDLSIAYTPGVASVSSAIAEDVELAYELT 60
M LGQLA+ QA+ GKGLEV KV ++ K DLSIAYTPGVASVSSAIA+D LAY+LT

60 Sbjct: 1 MKNQLGQLALEQAKTFGGKLEVPKVDIKTKHDLIAIYTPGVASVSSAIAKDKTLAYDLT 60

Query: 61 TKKNTVAVVSDGSAVLGLGDIGPEAAMPVMEGKAALFKRFANVDAVPIVLKTNDETEEIIIS 120
TKKNTVAV+SDG+AVLGLGDIGPEAAMPVMEGKAALFK FA VDA+PIVL T DTEEIIIS

Sbjct: 61 TKKNTVAVISDGTAVLGLGDIGPEAAMPVMEGKAALFKAFAGVDAIPIVLDTKDTTEEIIIS 120

-2540-

Query: 121 IVKAISPTFGGINLEDISAPRCFEIEQRLEEC DIPVFHDDQHGT AIVVLAALFNSLKL 180
 IVKA++PTFGGINLEDISAPRCFEIEQR LI+EC IPVFHDDQHGT AIVVLA A+FNSLKL+
 Sbjct: 121 IVKALAPTFGGINLEDISAPRCFEIEQR LIKECHIPVFHDDQHGT AIVVLA AIFNSLKL 180

Query: 181 KKDIEDIRVVVNGGGSAGLSITRKLLSAGAKHVTVD RFGI INDKDRESLAPHHKAI AKL 240
 KK +++++VVNGGGSAGLSITRKLL+AGA VTVVD+FGI IN+++ LAPHH IAK+
 Sbjct: 181 KKSLEVSIVVNGGGSAGLSITRKLLAAGATKVTVD KFGI INEQEAAQLAPHHLDIAKV 240

Query: 241 TNREFQSGSLEDALENADVF IGVSAP EALHA EWISKMA DKPIVFAMANPIE IYPDQALK 300
 TNREF+SG+LEDALE AD+FIGVSAP L AEWISKMA +P++FAMANPIE IYPD+AL+
 Sbjct: 241 TNREFKSGTLEDALEGADIF IGVSAP GVLKA EWISKMA ARPVI FAMANPIE IYPDEALE 300

Query: 301 AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIASLIPEEELS 360
 AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAA+GIASL+P++ LS
 Sbjct: 301 AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAKGIASLVPDDALS 360

Query: 361 TTHIIPNAFQNDVADV VAKSVSNAVQK 387
 TT+IIP+AF+ VA++VAKSV + V K
 Sbjct: 361 TTNIIIPDAFKEGVAEIVAKSVRSVVLK 387

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2249

A DNA sequence (GBSx2370) was identified in *S. agalactiae* <SEQ ID 6957> which encodes the amino acid sequence <SEQ ID 6958>. This protein is predicted to be Bta. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.02 Transmembrane 29 - 45 (29 - 45)

----- Final Results -----
 bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
 Identities = 35/112 (31%), Positives = 63/112 (56%)

Query: 1 MYSFEELLATMTLITAAETEDKIDSNQDFVLFIGRISCPFCHLFVPKIVEVADEDEFELF 60
 M F + + + + T + + D + FIGR +CP+C F + V E + + +
 Sbjct: 1 MEQFLDNIDKLEVTTVVRAQEALDKKETATFFIGRKTCPCYCRKFAGT LSGVVAETKAHIY 60

Query: 61 HLDSEDFDHW TANKEFRNKYDIPTVPGLMVVKNGTIKVKCD SKMTKEEIREF 112
 ++SE+ + FR++Y IPTVPG + + +G I V+CDS M+ +EI++F
 Sbjct: 61 FINSEEASQLNDLQAFRSRYGIPTVPGFVHITDGQINVRCDSSMSAQEI KDF 112

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6959> which encodes the amino acid sequence <SEQ ID 6960>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0900(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2541-

Identities = 39/111 (35%), Positives = 66/111 (59%)

Query: 3 SFEELLATMTLITAABIEDKIDSNDQDFVLFIGRISCPFCHLFVPKIVEVADEDEFELPHL 62
 +FEE++A + AE+ I S +D ++F+GR SCP+C F PK+ +VA +++ E++ +
 5 Sbjet: 11 TFEIIVANFIPSSVAEVTSAIASGKDMIVFLGRSSCPYCRRFAPKLAQVATDNQKEVYFV 70

Query: 63 DSEDFDHWITANKEFRNKYDIPTVPGLMVVKNGTIKVKCDSKMTKEBIREFI 113
 DSE+ FR Y + TVP L+V + + CDS +T ++I F+
 10 Sbjet: 71 DSENAADAELAAAFRENYQLVTVPALLVSYDQHQRVAVCDSSLTPDDILAF 121

SEQ ID 6958 (GBS427) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 5; MW 16.2kDa).

GBS427-His was purified as shown in Figure 214, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2250

A DNA sequence (GBSx2371) was identified in *S.agalactiae* <SEQ ID 6961> which encodes the amino acid sequence <SEQ ID 6962>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.75 Transmembrane 2 - 18 (1 - 21)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9437> which encodes amino acid sequence <SEQ ID 9438> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA11328 GB:D78257 ORF11 [Enterococcus faecalis]
 Identities = 36/80 (45%), Positives = 58/80 (72%)

Query: 1 MSLEPIIMLVVMVGMFFMQRQKKQAEQKQLNAVQKGDEIVTIGGLFGVVDEVNTEAQ 60
 M L +IML+V+V M F++ R QKKQ +ERQ LN +Q GD +VTIGGL GV+ E++++ +
 35 Sbjet: 1 MKLMLIMLVIVAMYFYLFRTQKKQKERQDFLNNLQPGDAVVTIGGLHGVISEISSDKK 60

Query: 61 RMVLDVDGVYLTFELAAIKS 80
 ++ LD +G + F+ +I++
 40 Sbjet: 61 KVTLDCEGAFFDFDQQSIRT 80

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6963> which encodes the amino acid sequence <SEQ ID 6964>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -6.10 Transmembrane 3 - 19 (1 - 22)
 INTEGRAL Likelihood = -3.03 Transmembrane 63 - 79 (63 - 79)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3442(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-2542-

>GP:BAA11328 GB:D78257 ORF11 [Enterococcus faecalis]
Identities = 29/75 (38%), Positives = 52/75 (68%)

5 Query: 6 ILMFVVMGLIWMQRRQKKQAQERQNLNAIEKGDEVVTIGGMFAIVDEVDTTAKKIVL 65
++M +V++ + +++ R QKKQ +ERQ+ LN ++ GD VVTIGG+ ++ E+ + KK+ L
Sbjct: 5 LIMLLVIVAMYFYLFRTQKKQKKERQDFLNNLQPGDAVVTIGGLHGVISEISSDKKKVTL 64

Query: 66 DVDGVFLTFELLAIK 80
D +G F F+ +I+

10 Sbjct: 65 DCEGAFDFDQQSIR 79

An alignment of the GAS and GBS proteins is shown below.

Identities = 63/90 (70%), Positives = 80/90 (88%)

15 Query: 4 PIIMLVVMVGMFFMQRRQKKQAQERQQLNAVQKGDEIVTIGGLFGVVDEVNTEAQRMV 63
PI+M VVM+G+++FMQRRQKKQAQERQ QLNA++KGDE+VTIGG+F +VDEV+T A+++V
Sbjct: 5 PILMFVVMGLIWMQRRQKKQAQERQNLNAIEKGDEVVTIGGMFAIVDEVDTTAKKIV 64

Query: 64 LDVDGVYLTFFELAAIKSVVSKAATPTEPVE 93
LDVDGV+LTFEL AIK +V+KA T T VE

20 Sbjct: 65 LDVDGVFLTFELLAIKRIVTKATTETTLVE 94

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2251

A DNA sequence (GBSx2372) was identified in *S.agalactiae* <SEQ ID 6965> which encodes the amino acid sequence <SEQ ID 6966>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have an uncleavable N-term signal seq

30 ----- Final Results -----
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 2252

A DNA sequence (GBSx2373) was identified in *S.agalactiae* <SEQ ID 6967> which encodes the amino acid sequence <SEQ ID 6968>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -1.38 Transmembrane 164 - 180 (164 - 180)

----- Final Results -----
bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB61731 GB:AL133220 putative oxidoreductase. [Streptomyces

-2543-

coelicolor A3(2)]
 Identities = 72/216 (33%), Positives = 120/216 (55%), Gaps = 1/216 (0%)

5 Query: 14 AQALEARGQKLYSVANRTYDKGLBFATKYGIQKVYDHIDQVFEDPEVDIIYISTPHNTHI 73
 A ++ ++ +VA+RT FA ++GI + Y + + D +VD++Y++TPH+ H
 Sbjct: 25 ADLVDLPDAEVVAVASRTEASAKTFAERFGIPRAYGGWETLARDEDVDVYVATPHSAHR 84

10 Query: 74 SFLRKALANGKHVLCESITLNSTELKEAIDLAEINHVVLAEAMTIFHMPYRQLKTLVD 133
 + L G++VLCEK TLN+ E E + LA N V L EAM ++ P+ R+LK LV
 Sbjct: 85 TAAGLCLEAGRNVLCEKPFITLNAREAAELVALARENGVFLMEAMMYCNPLVRRLKELVA 144

15 Query: 134 SGKLGPLKMIQMFNGSYKEYDMTNRFFSRDLAGGALLDIGVYALSCIRWFMSEAPHNITS 193
 G +G ++ +Q +FG + +R GGALLD+GVY +S + + E P ++ +
 Sbjct: 145 DGAIGEVRSLQADFLAGPFPAAHRLRDPAGGGGALLDLGVYPVSFAQLLLGE-PTDVAA 203

Query: 194 QVTFAPTGVDEQVGILLTNPANEMATVSLSLHAKQP 229
 + + GVD Q G LL+ + +A++ S+ P
 Sbjct: 204 RAVLSEEGVDLQTGALLSYGNDALASIHCSITGTP 239

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2253

A DNA sequence (GBSx2374) was identified in *S.agalactiae* <SEQ ID 6969> which encodes the amino acid sequence <SEQ ID 6970>. This protein is predicted to be surface protein Rib. Analysis of this protein
 25 sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4957(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2254

A DNA sequence (GBSx2375) was identified in *S.agalactiae* <SEQ ID 6971> which encodes the amino acid sequence <SEQ ID 6972>. This protein is predicted to be surface protein Rib. Analysis of this protein
 40 sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2544-

Example 2255

A DNA sequence (GBSx2376) was identified in *S.agalactiae* <SEQ ID 6973> which encodes the amino acid sequence <SEQ ID 6974>. This protein is predicted to be a host cell surface-exposed lipoprotein. Analysis of this protein sequence reveals the following:

```

5      Possible site: 38
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL    Likelihood = -7.75    Transmembrane    9 - 25 ( 5 - 28)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9005> which encodes amino acid sequence <SEQ ID 9006> was also identified. Analysis of this protein sequence reveals the following:

```

      Lipop: Possible site: -1    Crend: 3
      SRCFLG: 0
      McG: Length of UR: 24
          Peak Value of UR: 2.84
20      Net Charge of CR: 2
      McG: Discrim Score: 10.29
      GvH: Signal Score (-7.5): -4.34
          Possible site: 34
      >>> Seems to have an uncleavable N-term signal seq
25      Amino Acid Composition: calculated from 1
      ALOM program count: 1 value: -7.75 threshold: 0.0
          INTEGRAL    Likelihood = -7.75    Transmembrane    5 - 21 ( 1 - 24)
          PERIPHERAL Likelihood = 13.31      86
          modified ALOM score: 2.05
30      icml HYPID: 7 CFP: 0.410

      *** Reasoning Step: 3

      ----- Final Results -----
35      bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40      >GP:AAC03455 GB:AF020798 putative host cell surface-exposed
          lipoprotein [Streptococcus thermophilus bacteriophage TP-J34]
          Identities = 40/102 (39%), Positives = 63/102 (61%), Gaps = 10/102 (9%)

      Query: 101 KNALISAKIYSKTMNLSKQSIFEQLYSESPDKATHSDKFTKEESQYAIIDLKVDKFNAL 160
          + A+ AK Y+ T+++SK+ + QL S          DK++++ S YA+++ +D+ + AL
45      Sbjct: 51 RTAVSKAKQYASTVHMSKEELRSQLV-----FDKYSQDASDYAVENSGIDYNKQAL 102

      Query: 161 ETAKSYQSSSSLSKEEIIYKQLTSTLGDKFTNDEAQYAVDHLK 202
          E AK YQ + S+S + I QL S          DKFT +EA YAV +LK
50      Sbjct: 103 EKAKQYQDTLSMSPDAIRDQLVSF--DKFTQEADYAVANLK 142
          Identities = 40/112 (35%), Positives = 64/112 (56%), Gaps = 9/112 (8%)

      Query: 41 KKAKIKFNKTQKKIVKKAREYAKSGHMSKDSIIEKLKDSKKYRQEDINFVINLKVVDYK 100
          + ++ K K + V KA++YA + HMK+ + +L K Y Q+ ++ + N +DY
55      Sbjct: 40 QSSSEKVPKEYRTAVSKAKQYASTVHMSKEELRSQLVSF--FDKFTQEADYAVANLK 98

      Query: 101 KNALISAKIYSKTMNLSKQSIFEQLYSESPDKATHSDKFTKEESQYAIIDLK 152
          K AL AK Y T+++S +I +QL S          DKFT+EE+ YA+ +LK
60      Sbjct: 99 KQALEKAKQYQDTLSMSPDAIRDQLVS-----FDKFTQEADYAVANLK 142

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-2545-

SEQ ID 9006 (GBS122) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 6; MW 21.9kDa).

GBS122-His was purified as shown in Figure 202, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2256

A DNA sequence (GBSx2377) was identified in *S.agalactiae* <SEQ ID 6975> which encodes the amino acid sequence <SEQ ID 6976>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

```

10   Possible site: 42
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
15         bacterial cytoplasm --- Certainty=0.2830(Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
         bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20   >GP:CAB90833 GB:AJ250837 hypothetical protein [Streptococcus dysgalactiae]
    Identities = 91/96 (94%), Positives = 93/96 (96%)

    Query: 1  MSRKVRRHFTDDFKQQIVDLYNVGRKRSSLIKVYELTPSTFDKWVRQAKTTGSFKSIDNL 60
             MSRK+RRHFTDDFKQQIVDLYN GRKRSSLIK YELTPSTFDKWVRQAKTTGSFKS+DNL
    Sbjct: 1  MSRKIRRHFTDDFKQQIVDLYNAGRKRSSLIKEYELTPSTFDKWVRQAKTTGSFKSVDNL 60

25   Query: 61 TDEQRELIELRKHNKELEMQLDILKQAAVIMAQKGK 96
             TDEQRELIELRK NKELEMQLDILKQAAVIMAQKGK
    Sbjct: 61 TDEQRELIELRKRNKELEMQLDILKQAAVIMAQKGK 96

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2257

A DNA sequence (GBSx2378) was identified in *S.agalactiae* <SEQ ID 6977> which encodes the amino acid sequence <SEQ ID 6978>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

```

35   Possible site: 16
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40         bacterial cytoplasm --- Certainty=0.2618(Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
         bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9915> which encodes amino acid sequence <SEQ ID 9916> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9903> which encodes amino acid sequence <SEQ ID 9904> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2546-

>GP:CAB90834 GB:AJ250837 putative transposase [Streptococcus dysgalactiae]
Identities = 243/259 (93%), Positives = 250/259 (95%)

5 Query: 1 MCRWLNMPHSSYYYQAVESVSETFEETIKRIFLDSESRYGSRKIKICLNNEGITLSRRR 60
MCRWLN+P SSSYY+AVE VSE E EE+IK IFL+S++RYGSRKIKICLNNEGITLSRRR
Sbjct: 1 MCRWLNIPRSSYYYKAVEPVSEAELEESIKAFLESKARYGSRKIKICLNNEGITLSRRR 60

10 Query: 61 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDROFKQERPLQALVTDLTYYVRVGNR 120
IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDROFK ERPLQALVTDLTYYVRVGNR
Sbjct: 61 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDROFKPERPLQALVTDLTYYVRVGNR 120

15 Query: 121 WAYVCLIIDLYNREIIGLSLGHHTAELVKQAIQSIPYALTKVKMFHSDRGKEFDNQLID 180
WAYVCLIIDLYNREIIGLSLGHHTAELVKQAIQSIPY LTKVKMFHSDRGKEF+NQLID
Sbjct: 121 WAYVCLIIDLYNREIIGLSLGHHTAELVKQAIQSIPYPLTKVKMFHSDRGKEFNNQLID 180

Query: 181 EILEAFGITRSLSQAGCPYDनावेष्टYRAFKIEFVYQETPQLLEELALKTKDYVHWNNY 240
EILEAFGITRSLSQAGCPYDनावेष्टYRAFKIEFVYQETPQ LEELALKTK YVHWNNY
Sbjct: 181 EILEAFGITRSLSQAGCPYDनावेष्टYRAFKIEFVYQETPQSLEELALKTKAYVHWNNY 240

20 Query: 241 HRIHGSLNYQTPMTKRLIA 259
HRIHGSLNYQTPMTKRLIA
Sbjct: 241 HRIHGSLNYQTPMTKRLIA 259

There is also homology to SEQ ID 32.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2258

A DNA sequence (GBSx2379) was identified in *S.agalactiae* <SEQ ID 6979> which encodes the amino acid sequence <SEQ ID 6980>. This protein is predicted to be pXO1-128. Analysis of this protein sequence
30 reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3684(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAD32432 GB:AF065404 pXO1-128 [Bacillus anthracis]
Identities = 45/69 (65%), Positives = 52/69 (75%)

Query: 17 MKKAGKSNRVIMETLGIKNNSQIYTWMKWYENEELYRFHQGVGKQYTYGKGLEHLSEVEQ 76
MKK SNR IME LGIKN SQI TWMKWY ++ YRF Q VGKQY+YGKG + LSE+EQ
45 Sbjct: 1 MKKESYSNRTIMEKLGIKNVSIQKTWMKWYRTDQTYRFQGPVGKQYSYGKPKELSELEQ 60

Query: 77 LQLQVDLLK 85
L+L+ LK
50 Sbjct: 61 LRLENKHLK 69

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2547-

Example 2259

A DNA sequence (GBSx2380) was identified in *S.agalactiae* <SEQ ID 6981> which encodes the amino acid sequence <SEQ ID 6982>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 25
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2260

A DNA sequence (GBSx2382) was identified in *S.agalactiae* <SEQ ID 6985> which encodes the amino acid sequence <SEQ ID 6986>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```

    Possible site: 18
    >>> May be a lipoprotein

    ----- Final Results -----
25      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1595> which encodes the amino acid sequence <SEQ ID 1596>. Analysis of this protein sequence reveals the following:

```

    Possible site: 18
    >>> May be a lipoprotein

    ----- Final Results -----
35      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40   Identities = 302/306 (98%), Positives = 303/306 (98%)

   Query: 1   MKKVFFLMAMVSVSLVMIAGCDKSNPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVIRMIQ 60
           MKK FFLMAMVSVSLVMIAGCDKSNPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVIRMIQ
   Sbjct: 1   MKKGFFLMAMVSVSLVMIAGCDKSNPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVIRMIQ 60

45   Query: 61  SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSQVNVFEASKPLTLDR 120
           SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSQVNVFEASKPLTLDR
   Sbjct: 61  SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSQVNVFEASKPLTLDR 120

50   Query: 121 VKGLEDMETVQGIDPATLYDPHTWTDPLVAGEEAVNIAKELGHLDPKHKDSYTKKAKAFK 180
           VKGLEDMETVQGIDPATLYDPHTWTDPLVAGEEAVNIAKELG LDPKHKDSYTK AKAFK
   Sbjct: 121 VKGLEDMETVQGIDPATLYDPHTWTDPLVAGEEAVNIAKELGRLDPEKHKDSYTKNAKAFK 180

   Query: 181 KEAEQLTEBYTQKFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISPEQEPSRQLKEI 240

```

-2548-

```

      KEAEQLTTEEYTKQFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISEQEPSRQLKEI
Sbjct: 181 KEAEQLTTEEYTKQFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISEQEPSRQLKEI 240

      QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV 300
5      QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV
      Sbjct: 241 QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV 300

      Query: 301 LYQQLK 306
      LYQQLK
10     Sbjct: 301 LYQQLK 306

```

There is also homology to SEQ ID 4.

SEQ ID 6986 (GBS189) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 2; MW 35.2kDa).

15 The GBS189-His fusion product was purified (Figure 204, lane 7) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 248A), FACS (Figure 248B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 2261

A DNA sequence (GBSx2383) was identified in *S.agalactiae* <SEQ ID 6987> which encodes the amino acid sequence <SEQ ID 6988>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4656(Affirmative) < succ>
30      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB41455 GB:U34956 phosphoribosylformylglycinamide synthase
[Mycobacterium tuberculosis]
35 Identities = 73/237 (30%), Positives = 112/237 (46%), Gaps = 25/237 (10%)

      Query: 43 GAGGVCVAIGELAD---GLEIDLKVKPLKYQGNGTBIATSESQERMSVVVGPSDVDAF 98
      G G+ A ELA G+ I LD VPL+ + + E+ SESQERM VV P +VDAF
      Sbjct: 282 GGAGLSCATSELASAGDGGMTIQLDSVPLRAKEMTPABVLCSESQERMCAVVSEKQNVDAF 341

40      Query: 99 IAACNKENIDAVVVATVTEKPNLVMVTWNGETIVDLERCFLDTNG-----VRVVVDAKVV 152
      +A C K + A V+ VT+ L +TW+GET+VD+ + G V +
      Sbjct: 342 LAVCRKWEVLATVIGEVTDGDRQLQITWHGETVVDVPPRTVAHEGVPVYQRPVARPDTQDAL 401

45      Query: 153 DKDLTVPEARTTSAETLEADMLKVLSDLNHASQKGLQTIFDSSVGRSTV--NHPIGGRYQ 210
      + D + +R + + L A +L +L + S+ + +D V +TV H GG +
      Sbjct: 402 NADRSAKLSRPVTGDELRAITLALLGSPHLCSTRAFIQYDRYVRGNTVLAEHADGGMLR 461

      Query: 211 ITPTESSVQKLPVQYGVTTTASVMAQGYNPYIAEWSPPYHGAAYAVIEATARLVATGA 267
50      I ES+ + + V + +++ PY GA A+ EA + TGA
      Sbjct: 462 I--DESTGRGIADVSTDASGRYTLL-----DPYAGQLALAEAYRNVAVTGA 505

```

There is also homology to SEQ ID 982.

-2549-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2262

A DNA sequence (GBSx2384) was identified in *S.agalactiae* <SEQ ID 6989> which encodes the amino acid sequence <SEQ ID 6990>. This protein is predicted to be 30S ribosomal protein S11 (rpsK). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0598(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9281> which encodes amino acid sequence <SEQ ID 9282> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10919> which encodes amino acid sequence <SEQ ID 10920> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11918 GB:Z99104 ribosomal protein S11 (BS11) [Bacillus subtilis]
Identities = 81/92 (88%), Positives = 87/92 (94%)

Query: 2 HGNALAWSSAGALGFKGSRKSTPFAAQMAEAAKSAQEHGLKTVEVTVKGP GSGRESAI 61
HGNA++WSSAGALGF+GSRKSTPFAAQMAE AAK + EHGLKT+EVTVKGP GSGRE+AI
Sbjct: 40 HGNALSWSSAGALGFRGSRKSTPFAAQMAETAAKGSI EHGLKTLEVTVKGP GSGREAAI 99

Query: 62 RALAAAGLEVTAI RDVTPVPHNGARPPKRRRV 93
RAL AAGLEVTAI RDVTPVPHNG RPPKRRRV
Sbjct: 100 RALQAAGLEVTAI RDVTPVPHNGCRPPKRRRV 131

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6991> which encodes the amino acid sequence <SEQ ID 6992>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0945(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 92/93 (98%), Positives = 93/93 (99%)

Query: 1 MHGNALAWSSAGALGFKGSRKSTPFAAQMAEAAKSAQEHGLKTVEVTVKGP GSGRESA 60
+HGNALAWSSAGALGFKGSRKSTPFAAQMAEAAKSAQEHGLKTVEVTVKGP GSGRESA
Sbjct: 35 VHGNALAWSSAGALGFKGSRKSTPFAAQMAEAAKSAQEHGLKTVEVTVKGP GSGRESA 94

Query: 61 IRLAAAGLEVTAI RDVTPVPHNGARPPKRRRV 93
IRLAAAGLEVTAI RDVTPVPHNGARPPKRRRV
Sbjct: 95 IRLAAAGLEVTAI RDVTPVPHNGARPPKRRRV 127

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2550-

Example 2263

A DNA sequence (GBSx2385) was identified in *S.galactiae* <SEQ ID 6993> which encodes the amino acid sequence <SEQ ID 6994>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2551(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB03881 GB:AP001507 DNA-directed RNA polymerase alpha subunit
 [Bacillus halodurans]

Identities = 190/314 (60%), Positives = 249/314 (78%), Gaps = 2/314 (0%)

Query: 1 MIEFEKPIITKIDENKD--YGRFVIEPLERGYGTTLGNSLRRLSSLPAAVTSIKIDG 58
 MIE EKP+I I+ ++D YG+FV+EPLERGYGTTLGNSLRRLSSLPAAVTS++IDG
 Sbjct: 1 MIEIEKPVIEITIEISEDAKYGKFVVEPLERGYGTTLGNSLRRLSSLPAAVTSVQIDG 60

Query: 59 VLHEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVEGPAEITAGDILTDSIDIEI 118
 VLHEF TI GV EDV I+LN+K LA+K Y +++K +E+D +G +TAGD+ DSD+++
 Sbjct: 61 VLHEFSTIEGVVEDVTITVLNLKQLALKIYSDDEKTLTIDTQEGGVVTAGDLTHDSVDV 120

Query: 119 VNPFDHYLFTIAEGHSLKATMTVAKNRGYVPAEGNKDDAPVGTAVDSIYTPVKKVNYQV 178
 +NED ++ T+ G L+ +T + RGYVPAEGNK D+ +G + +DSIYTPV +VNYQV
 Sbjct: 121 LNPDLHIATLTGHAHRMITAKRGRGYVPAEGNKSDLAIGVIPIDSITPVSRVNYQV 180

Query: 179 EPARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLEHLNLTDLTEVAKATEVMKET 238
 E RVG +DKLT+++ T+G+I PE+A+ L A++L EHLN+F LT+ A+ E+M E
 Sbjct: 181 ENTRVGQVTNYDKLTLDVWTDGSIREEAVSLGAKILTEHLNIFVGLTDQAQNAEIMVEK 240

Query: 239 EKVNDKVLDRITIEELDLSVRSYNCLKRAGINTVFDLTEKTEPEMMKVRNLGRKSLEEVK 298
 E+ EKVL+ TIEELDLSVRSYNCLKRAGINTV +LT+KTE +MMKVRNLGRKSLEEV+
 Sbjct: 241 EEDQKEKVLMTIEELDLSVRSYNCLKRAGINTVQELTQKTEEDMMKVRNLGRKSLEEVQ 300

Query: 299 IKLADLGLGLKNDK 312
 KL +LGLGL+ ++
 Sbjct: 301 EKLGLGLGLRKEE 314

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6995> which encodes the amino acid sequence <SEQ ID 6996>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2551(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 305/312 (97%), Positives = 311/312 (98%)

Query: 1 MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRLSSLPAAVTSIKIDGVL 60
 MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRLSSLPAAVTSIKIDGVL
 Sbjct: 1 MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRLSSLPAAVTSIKIDGVL 60

Query: 61 HEFDITIPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVEGPAEITAGDILTDSIDIEIV 120
 HEFDITIPGVREDVMQIILNVKGLAVKSYVEDEKIIEL+VEGPAE+TAGDILTDSIE+VN
 Sbjct: 61 HEFDITIPGVREDVMQIILNVKGLAVKSYVEDEKIIELEVEGPAEVTAGDILTDSIDIELVN 120

-2551-

Query: 121 PDHYLFTIAEGHSLKATMTVAKNRGYVPAEGNKKDDAPVGT LAVDSIYTPVKVNYQVEP 180
 PDHYLFTIAEGHSL+ATMTVAK RGYVPAEGNKKDDAPVGT LAVDSIYTPVKVNYQVEP
 Sbjct: 121 PDHYLFTIAEGHSLRATMTVAKKRGYVPAEGNKKDDAPVGT LAVDSIYTPVKVNYQVEP 180

5 Query: 181 ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLI EHLNLFDTL TEVAKATEVMKETEK 240
 ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLI EHLNLFDTL TEVAKATEVMKETEK
 Sbjct: 181 ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLI EHLNLFDTL TEVAKATEVMKETEK 240

10 Query: 241 VNDEKVLDR TIEELDLSVRSYNCLK RAGINTVFDL TEKTEPEMMKVRNLGRKSLEEVKIK 300
 VNDEKVLDR TIEELDLSVRSYNCLK RAGINTVFDL TEK+EP EMMKVRNLGRKSLEEVK+K
 Sbjct: 241 VNDEKVLDR TIEELDLSVRSYNCLK RAGINTVFDL TEKSEPEMMKVRNLGRKSLEEVKVK 300

Query: 301 LADLGLGLKNDK 312
 LADLGLGLKNDK
 15 Sbjct: 301 LADLGLGLKNDK 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2264

20 A DNA sequence (GBSx2386) was identified in *Sagalactiae* <SEQ ID 6997> which encodes the amino acid sequence <SEQ ID 6998>. This protein is predicted to be 50S ribosomal protein L17 (rp1Q). Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1609(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11920 GB:Z99104 ribosomal protein L17 (BL15) [Bacillus subtilis]
 Identities = 95/128 (74%), Positives = 105/128 (81%), Gaps = 8/128 (6%)

35 Query: 1 MAYRKLGR TSSQRKAMLRDLITD L LINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR 60
 M+YRKLGR TS+QRKAMLRDLITDL+INE I TTE RAKE+R VEKMITLGKRGDLHARR
 Sbjct: 1 MSYRKLGR TSAQRKAMLRDLITD LIINERIETTETRAKELRSVVEKMITLGKRGDLHARR 60

40 Query: 61 QAAAYVRNEIASENYDEASDKYTSTTALQKLFDDIAPRYAERNNGGYTRILKTEPRRGDAA 120
 QAAAY+RNE+A+E ++ ALQKLF DIA RY ER GGYTRI+K PRRGD A
 Sbjct: 61 QAAAYIRNEVANEENNQ-----DALQKLFSDIATRYEERQGGYTRIMKLGPRRGDGA 112

Query: 121 PMAIIELV 128
 PMAIIELV
 45 Sbjct: 113 PMAIIELV 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6999> which encodes the amino acid sequence <SEQ ID 7000>. Analysis of this protein sequence reveals the following:

Possible site: 37
 50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1609(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/128 (97%), Positives = 127/128 (98%)

-2552-

Query: 1 MAYRKLGRSSQKAMLRDLTTDLLINESIVTTEARAKEIRKTVKMITLGKRGDLHARR 60
 MAYRKLGRSSQKAMLRDLTTDLLINESIVTTEARAKEIRKTVKMITLGKRGDLHARR
 5 Sbjct: 1 MAYRKLGRSSQKAMLRDLTTDLLINESIVTTEARAKEIRKTVKMITLGKRGDLHARR 60

Query: 61 QAAAYVRNEIASENYDEASDKYTSTTALQKLFDDIAPRYAERNNGGYTRILKTEPRRGDAA 120
 QAAAYVRNEIASENYDEA+DKYTSTTALQKLF +IAPRYAERNNGGYTRILKTEPRRGDAA
 Sbjct: 61 QAAAYVRNEIASENYDEATDKYTSTTALQKLFSEIAPRYAERNNGGYTRILKTEPRRGDAA 120

10 Query: 121 PMAIIELV 128
 PMAIIELV
 Sbjct: 121 PMAIIELV 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 15 vaccines or diagnostics.

Example 2265

A DNA sequence (GBSx2396) was identified in *S.agalactiae* <SEQ ID 7001> which encodes the amino acid sequence <SEQ ID 7002>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

20 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAA83977 GB:AF138877 mercuric reductase MerA [Bacillus sp.
 RC607]
 Identities = 29/33 (87%), Positives = 32/33 (96%)

Query: 4 VGLTEEQAKEKGYDVKTSVLPLXAVPRAIVNRE 36
 VGLTE+QAKEKGY+VKTSVLPL AVPRA+VNRE
 35 Sbjct: 520 VGLTEQQAKEKGYEVKTSVLPLDAVPRALVNRE 552

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 2266

A DNA sequence (GBSx2397) was identified in *S.agalactiae* <SEQ ID 7003> which encodes the amino acid sequence <SEQ ID 7004>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

45 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3016(Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]

-2553-

Identities = 146/194 (75%), Positives = 175/194 (89%)

Query: 2 PQISGLEKMDYLTSTTLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQSRERLL 61
 P I GL ++DYLTST+LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QSRERLL
 5 Sbjct: 226 PNIPGLNEVDYLTSTLLELKKVPKRLVVIGSGYIGMELGQLFHHLGSEVTLIQSRERLL 285

Query: 62 KEYDPEISESVEKALIEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVAT 121
 KEYDPEISESVEK+L+EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVAT
 10 Sbjct: 286 KEYDPEISESVEKSLVEQGGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVAT 345

Query: 122 GRKPNTDSLNL SAAGVETGKNNELINDFGQTSNEKIYAAGDVTLGPFQFVYVAAYEGGII 181
 GR PNT +LNL AAGVE G EI+I+D+ +T+N +IYAAGDVTLGPFQFVYVAAY+GG+
 15 Sbjct: 346 GRTPNTATLNLRAAGVEIGSRGEIIDDYSRTTNTRIYAAGDVTLGPFQFVYVAAYQGGVA 405

Query: 182 TDNAIGGLNKKIDL 195
 NAIGGLNKK++L
 15 Sbjct: 406 APNAIGGLNKKLNL 419

There is also homology to SEQ ID 1820.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2267

A DNA sequence (GBSx2398) was identified in *S.galactiae* <SEQ ID 7005> which encodes the amino acid sequence <SEQ ID 7006>. This protein is predicted to be triacylglycerol acylhydrolase. Analysis of this
 25 protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3180(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2268

A DNA sequence (GBSx2399) was identified in *S.galactiae* <SEQ ID 7007> which encodes the amino acid sequence <SEQ ID 7008>. Analysis of this protein sequence reveals the following:
 40

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0544(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAC74453 GB:AE000234 orf, hypothetical protein [Escherichia coli K12]
 Identities = 45/58 (77%), Positives = 51/58 (87%)

-2554-

Query: 1 MPWQNLLHAGQENLFSGLTALTAEFTVGEGLMTHDEPCSMAPDDKHDLSIGTCSHLP 58
 +PWQNLLHAG+ENLFSGLTAL+AEFT+GEG+LM HD P APD+ DLISGTCSHLP
 Sbjct: 34 LPWQNLLHAGEENLFSGLTALSABFTIGEGELMAHDVPLGCAPDEYDDLISGTCSHLP 91

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2269

10 A DNA sequence (GBSx2400) was identified in *S.agalactiae* <SEQ ID 7009> which encodes the amino acid sequence <SEQ ID 7010>. This protein is predicted to be transposase for insertion sequence element is5. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2058(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB15497 GB:AK026530 unnamed protein product [Homo sapiens]
 Identities = 297/299 (99%), Positives = 297/299 (99%)

25 Query: 1 MEQILFWQNMVEVIEPFYKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM 60
 MEQILFWQNMVEVIEPFYKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM
 Sbjct: 40 MEQILFWQNMVEVIEPFYKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM 99

30 Query: 61 RLFARLSLDSALPDRTTIMNFRHLLBQHQLARQLFKTINRWLAEGVMMTQGTTLVDATII 120
 RLFARLSLDSALPDRTTIMNFRHLLBQHQLARQLFKTINRWLAEGVMMTQGTTLVDATII
 Sbjct: 100 RLFARLSLDSALPDRTTIMNFRHLLBQHQLARQLFKTINRWLAEGVMMTQGTTLVDATII 159

35 Query: 121 EAPSSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTSLVTTAANEHDLNQLX 180
 EAPSSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTSLVTTAANEHDLNQL
 Sbjct: 160 EAPSSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTSLVTTAANEHDLNQLG 219

40 Query: 181 NLLHGEEQFVSADAXYQGAPQREELA EVDVDWLIAERPGKVRTLKQHPKKNKTAINIEYM 240
 NLLHGEEQFVSADA YQGAPQREELA EVDVDWLIAERPGKVRTLKQHPKKNKTAINIEYM
 Sbjct: 220 NLLHGEEQFVSADAGYQGAPQREELA EVDVDWLIAERPGKVRTLKQHPKKNKTAINIEYM 279

40 Query: 241 KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH 299
 KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH
 Sbjct: 280 KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH 338

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2270

A DNA sequence (GBSx2401) was identified in *S.agalactiae* <SEQ ID 7011> which encodes the amino acid sequence <SEQ ID 7012>. Analysis of this protein sequence reveals the following:

50 Possible site: 16
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>

-2555-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5

>GP:CAB51958 GB:AL109661 putative eukaryotic-type serine/threonine
 protein kinase [Streptomyces coelicolor A3(2)]
 Identities = 49/169 (28%), Positives = 90/169 (52%), Gaps = 6/169 (3%)

10

Query: 23 PTTIRVPDVSNKTVAQAKMTLENSGLKVGAI RNIESDSVSEGLVVKTDPAAGRSRREGAK 82
 P T+++PDV+ + +A+ LE+ GL+ G + SD V+ G V+ T P +G + R G+
 Sbjct: 469 PDTVKLPDVTGYKLDKARTLLEDEGLEPGMVTRAFSDEVARGFVISTKPGSGTTVRAGSA 528

15

Query: 83 VNLYIATPNKSFTLGNYKEHNYKDILKDLQKGVKKS LKVKRKINNDYTTGTILAQSLP 142
 V L + + + + + + + +L+G G+K + ++N++Y +G + A+ P
 Sbjct: 529 VAL-VVSKGSPVDVPDVTGDDLDEARAELEGAGLK--VKTADERVNSEYDSGRV-ARQTP 584

20

Query: 143 EGTSEFNDGNNKLTTLTVAVNDPMI-MPDVTGMTVGEVIETLTDLGLDAD 190
 E +G+ +TLTV+ MI +PDV G +V + + L D G + D
 Sbjct: 585 EPGGRAAEGD-TVTLTVSKGPRMIEVPDVVGDSVDDAKQKLEDAGFEVD 632
 Identities = 45/161 (27%), Positives = 80/161 (48%), Gaps = 4/161 (2%)

25

Query: 27 RVPDVSNKTVAQAKMTLENSGLKVGAI RNIESDSVSEGLVVKTDPAAGRSRREGAKVNLY 86
 +VP + +KT AQA+ L+++GL VG +R+ SD+V G V+ TDP G R+ V+L
 Sbjct: 405 KVPPLLSKTEAQARDRLDDAGLDVGKVRHAYS DTVERGKVISTDPGVGDRIRKNDVSLT 464

30

Query: 87 IATPNKSFTLGNYKEHNYKDILKDLQKGVKKS LKVKRKINNDYTTGTILAQSLPEGTS 146
 ++ + L + + + L+ +G++ + V R +++ G +++ GT+
 Sbjct: 465 VSDGPDVTVKLPDVTGYKLDKARTLLEDEGLEPGM--VTRAFSDEVARGFVISTKPGSGTT 522

Query: 147 FNPDGNKLTTLTVAVNDEMIMPDVTGMTVGEVIETLTDLGL 187
 + L V+ P+ +PDVTG + E L GL
 Sbjct: 523 VR--AGSAVALVSKGSPVDVPDVTGDDLDEARAELEGAGL 561

35 There is also homology to SEQ ID 3026.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2271

40 A DNA sequence (GBSx2402) was identified in *S. agalactiae* <SEQ ID 7013> which encodes the amino acid sequence <SEQ ID 7014>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq

45

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50

A related GBS nucleic acid sequence <SEQ ID 9311> which encodes amino acid sequence <SEQ ID 9312> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55

>GP:AAB90561 GB:AE001058 glutamine ABC transporter, ATP-binding
 protein (glnQ) [Archaeoglobus fulgidus]
 Identities = 142/219 (64%), Positives = 178/219 (80%)
 Query: 1 MDIHQGEVVVVIIGPSGSGKSTFLRTMNLLEVPTKGTVTTFEGIDITDKKNDIFKMREKMG 60
 M + +GEVVVVIIGPSGSGKST LR +N LE PT G + +G+DIT+ K DI K+R+++G+
 Sbjct: 24 MKVEKGEVVVVIIGPSGSGKSTLLRCINRLEPTSGKILLDGVDTNSKIDINKVRQRIGI 83

-2556-

Query: 61 VFQQFNLFNMTVLENITLSPIKTKGLSNLDAQTKAYELLEKVGLKEKANTYPASLSGGQ 120
 VFOQFNLFPP++T L+N+TL+PIK K +S +A+ LLEKVGL++KA+ YPA LSGGQ
 Sbjct: 84 VFQQFNLFPHLTALQNVTLAPIKIKKMSKREAEELGMRLLEKVGLLEDKADYYPQLSGGQ 143

Query: 121 QQRIAIARGLAMNPVLLFDEPTSAIDPEMVGEVLTVMQDLAKSGMTMVIVTHEMGFARE 180
 QQR+AIAR LAMNP+V+LFDE TSALDPE+V EVL VM+ LA+ GMTMV+VTHEMGFARE
 Sbjct: 144 QQRVAIARALAMNPEVMLFDEVTSAIDPELVKEVLDVMKQLARDGMTMVVVTHEMGFARE 203

Query: 181 VADRVIFMDAGIIVEQGAPKEVFEQTKEIRTRDFLSKVL 219
 V DRVIFMD G+IVE+G P+++F K RTR FLS +L
 Sbjct: 204 VGDRVIFMDGGVIVEEGKPEQIFSNPKHERTRKFLSMIL 242

There is also homology to SEQ ID 1186.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2272

A DNA sequence (GBSx2403) was identified in *S.agalactiae* <SEQ ID 7015> which encodes the amino acid sequence <SEQ ID 7016>. This protein is predicted to be 4-hydroxy-2-oxoglutarate aldolase (kdgA).

- Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1479(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14127 GB:Z99115 deoxyphosphogluconate aldolase [Bacillus subtilis]
 Identities = 21/62 (33%), Positives = 38/62 (60%), Gaps = 4/62 (6%)

Query: 3 QLMQKIVAVIRGNSQEEAFQAAQACIKGGISAIEIAYTNSKASQVIEQLVTQYTNQEQV 62
 +L + K++AVIR ++EA Q ++ + GI A+E+ YT AS +IE + N+E +
 Sbjct: 9 RLKEAKLIAVIRSKDKQEACQQIESLLDKGIRAVEVTTYTPGASDIIE----SFRNREDI 64

Query: 63 VV 64
 ++
 Sbjct: 65 LI 66

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2273

- A DNA sequence (GBSx2405) was identified in *S.agalactiae* <SEQ ID 7017> which encodes the amino acid sequence <SEQ ID 7018>. This protein is predicted to be H repeat-associated protein (rfbQRS) (b1458). Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0207(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is homology to SEQ ID 504.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2274

- 5 A DNA sequence (GBSx2406) was identified in *S.agalactiae* <SEQ ID 7019> which encodes the amino acid sequence <SEQ ID 7020>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -6.74    Transmembrane    2 - 18 ( 1 - 21)
10    INTEGRAL    Likelihood = -3.03    Transmembrane    73 - 89 ( 73 - 92)

----- Final Results -----
                bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 3376.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2275

A DNA sequence (GBSx2407) was identified in *S.agalactiae* <SEQ ID 7021> which encodes the amino acid sequence <SEQ ID 7022>. This protein is predicted to be insertion element IS1 protein InsB (insB_5). Analysis of this protein sequence reveals the following:

```

25 Possible site: 52
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.4280(Affirmative) < succ>
30                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2276

A DNA sequence (GBSx2409) was identified in *S.agalactiae* <SEQ ID 7023> which encodes the amino acid sequence <SEQ ID 7024>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 13
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.3937(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

-2558-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2277

- 5 A DNA sequence (GBSx2410) was identified in *S.agalactiae* <SEQ ID 7025> which encodes the amino acid sequence <SEQ ID 7026>. This protein is predicted to be triosephosphate isomerase (tpi). Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence
10  INTEGRAL    Likelihood = -0.37    Transmembrane    35 - 51 ( 35 - 51)

----- Final Results -----
                bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC43268 GB:U07640 triosephosphate isomerase [Lactococcus
                lactis]
20  Identities = 50/75 (66%), Positives = 61/75 (80%)

Query: 6  IAGNWKMNKNPPEEAKAFIEAVASKLPSSSELVEAGIAAPALTLSTVLEAAKGSELKIAAQN 65
        IAGNWKMNK  EA+AF+EAV + LPSS+ VE+ I APAL L+ +      +GSELK+AA+N
Sbjct: 7  IAGNWKMNKTLSEAQAFVEAVKNNLPSSDNVESVIGAPALFLAPMAYLRQSELKLAEN 66

25  Query: 66 SYFENSGAFTGENSP 80
        SYFEN+GAFTGENSP
Sbjct: 67 SYFENAGAFTGENSP 81

```

- 30 There is also homology to SEQ ID 6838:.

```

Identities = 58/77 (75%), Positives = 68/77 (87%)

Query: 6  IAGNWKMNKNPPEEAKAFIEAVASKLPSSSELVEAGIAAPALTLSTVLEAAKGSELKIAAQN 65
        IAGNWKMNKNP+EAKAF+EAVASKLPS++LV+  +AAPA+ L T +EAAK S LK+AAQN
35  Sbjct: 7  IAGNWKMNKNPQEAKAFVEAVASKLPSTDLVDVAAPAVDLVTTIEAAKDSVLKVAQN 66

Query: 66 SYFENSGAFTGENSPKV 82
        YFEN+CAFTGE SPKV
Sbjct: 67 CYFENTGAFTGETSPKV 83
40

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2278

- 45 A DNA sequence (GBSx2412) was identified in *S.agalactiae* <SEQ ID 7027> which encodes the amino acid sequence <SEQ ID 7028>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -2.39    Transmembrane    96 - 112 ( 96 - 112)

50  ----- Final Results -----
                bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```


-2559-

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:BAA14368 GB:D90354 surface protein antigen precursor
    [Streptococcus sobrinus]
    Identities = 60/129 (46%), Positives = 76/129 (58%), Gaps = 18/129 (13%)

Query: 3  ISFDNSFLETVSDDSAFQADVYLQMKRIAAGQVENTYLHTVNGYVISSNTVVTHTPQPPE 62
      ++F  FL +VS DSAFQA+VYLQMKRIA G  NTY++TVNG  SSNTV T TP+P++
Sbjct: 1442 VTFKEDFLRSVSVDSAFQAEVYLQMKRIAVGTFANTYVNTVNGITYSSNTVRTSTPEPKQ 1501

10 Query: 63  PSPNPQ-----TPPQPPIETIEPPVPASILPNTGEQES----LLGLIG--AGILLGT 108
      PSP P          P Q          PP A LP TG+  +  LLGL+  AG L
Sbjct: 1502 PSPVDPKTTTTVVFQPRQGKAYQPAPPAGAQ-LPATGDSSNAYLPLLGLVSLTAGFSL-- 1558

15 Query: 109  AYGLKKKEE 117
      GL++K++
Sbjct: 1559 -LGLRRKQD 1566

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2279

A DNA sequence (GBSx2413) was identified in *S.agalactiae* <SEQ ID 7029> which encodes the amino acid sequence <SEQ ID 7030>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 23
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3691(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9359> which encodes amino acid sequence <SEQ ID 9360> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

35 >GP:CAB15793 GB:Z99123 phosphotransacetylase [Bacillus subtilis]
    Identities = 131/221 (59%), Positives = 169/221 (76%), Gaps = 2/221 (0%)

Query: 6  LVDPVILGKADEVHDSLARLGFVDQDYSIIDPEQYEFEEEMKEAFVEIRK GKATMEDADR 65
      +++P+++G  +E+  L          I DP YE  E++ +AFVE RKGKAT E A +
40 Sbjct: 41 VINPIVIGNENEIQAKAKELNLTGGVKTYDPHTYEGMEDLVQAFVERRK GKATEEQARK 100

Query: 66  LLKDVNYFGVMLVKLGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLMRENT 125
      L D NYFG MLV  GLADG+VSGA HSTADTVRPALQIIKTK G+ +TSGVF+M R
Sbjct: 101 ALLDENYFGTMLVYKGLADGLVSGAAHSTADTVRPALQIIKTKGKVKTSGVFMARG-- 158

45 Query: 126 QERYIFADCAINIDPNAQELAEIAVNTADTAKIFDIDPKIAMLSFSTKGSAPQAEKVQ 185
      +E+Y+PADCAINI P++Q+LAEIA+ +A+TAK+FDI+P++AMLSFSTKGSAP+ + EKV
Sbjct: 159 EEQYVFADCAINIAPDSQDLAEIAIESANTAKMFDIEPRVAMLSFSTKGSAPKSDTEKVA 218

50 Query: 186 EAAKIAKDLSPELAVDGELOFDAAFVPETAEIKAPNSDVAG 226
      +A KIAK+ +PEL +DGE QFDAAFVP AE KAP+S++ G
Sbjct: 219 DAVKIAKEKAPELTLDGEFQFDAAFVPSVAEKKAPDSEIKG 259

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7031> which encodes the amino acid sequence <SEQ ID 7032>. Analysis of this protein sequence reveals the following:

Possible site: 34

-2560-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3182(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 181/227 (79%), Positives = 211/227 (92%)

Query: 1 MKFEGLVDPVILGKADEVHDSLARLGFVDQDYSIIDPEQYKFEEMKEAFVEIRKKGKATM 60
 +KFEGL++P+ILG+++EV + L +LGF DQDY+II+P +Y F++MKEAFVE+RKGKAT+
 Sbjct: 38 LKFEGLLEPIILGQSEEVNLLTKLGFADQDYTIINPNEYADFDKMKEAFVEVRKKGKATL 97

15 Query: 61 EDADRLLLKDVNYFGVMLVKLGADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM 120
 EDAD++L+DVNYFGVMLVK+GLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM
 Sbjct: 98 EDADKMLRDVNYFGVMLVKMGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM 157

20 Query: 121 NRENTQERYIFADCAINIDPNAQELAEIAVNTADTAKIFDIDPKIAMLSFSTKGSAPQ 180
 NRENT ERY+FADCAINIDP AQELAEIAVNTA+TAKIFDIDPKIAMLSFSTKGS KAPQ
 Sbjct: 158 NRENTSERYVFADCAINIDPTAQELAEIAVNTAETAKIFDIDPKIAMLSFSTKGSKAPQ 217

 Query: 181 AEKVQEAAKIAKDLSPELAVDGELQFDAAFVPETAETAIKAPNSDVAGK 227
 +KV+EA +IA L+P+LA+DGELQFDAAFVPETA IKAP+S VAG+
 25 Sbjct: 218 VDKVREATEIATGLNPDALDGELOFDAAFVPETAETAIKAPDSAVAGQ 264

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2280

30 A DNA sequence (GBSx2414) was identified in *S.agalactiae* <SEQ ID 7033> which encodes the amino acid sequence <SEQ ID 7034>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.4076(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG19110 GB:AE005009 Vng0600c [Halobacterium sp. NRC-1]
 Identities = 57/176 (32%), Positives = 86/176 (48%), Gaps = 20/176 (11%)

45 Query: 1 MKVLLYLEAEYELKKSIGRAIKHQEKALQIAGIDYTTNPT----- 41
 M+ L YLEA E L+ G+ A Q AL+ ++ P
 Sbjct: 2 MRALNYLEAAEALR-GGMVTATNQRAALETVDVEVVETPWRAGDPVRSIGSLAAGGSCF 60

50 Query: 42 DDFDLVHMNTYGIRSWLLMSKAKKTGKKVIMHGHSTEDFRNSFIGSNLVSELPFKWYLCR 101
 FD+ H N G S + A++T +++H H T EDF SF GS+ ++P + YL
 Sbjct: 61 TAFDVAHCNLVGPGSVAVARHARRITDPLVLHAHLTREDFAQSFRGSSTIAPALEPYLRW 120

 Query: 102 FYQKADAIITPTDYSKQLIKAYGIKKPIFVLSNGIDLSRYQXSEKKESAFRHYFHL 157
 FY +AD ++ P++Y+K +++AY + PI LSNG+DL Q E + R F L
 55 Sbjct: 121 FYSQADLVLCPSYTKDVLRAYPVDAPIRQLSNGVDLESMQGYESFRADTRARFDL 176

There is also homology to SEQ ID 1220.

-2561-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2281

A DNA sequence (GBSx2415) was identified in *S.agalactiae* <SEQ ID 7035> which encodes the amino acid sequence <SEQ ID 7036>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2625(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC35010 GB:AF055987 intracellular a-amylase [Streptococcus mutans]
 Identities = 27/46 (58%), Positives = 33/46 (71%)

Query: 1 MEVGEIYAGKTFVDYLGNCQEYVIGDDGWGDFLVESASISAWVPK 46
 M +GE K FVDYL NC +EV++ D GWGDF V+ AS+SAWV K
20 Sbjct: 438 MNMGEFNRNKVFVDYLNNTCTEEVILDDQGWGDFPVQEASLSAWVNK 483

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2282

A DNA sequence (GBSx2416) was identified in *S.agalactiae* <SEQ ID 7037> which encodes the amino acid sequence <SEQ ID 7038>. This protein is predicted to be RopA. Analysis of this protein sequence reveals the following:

Possible site: 24
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2082(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 6908:

Identities = 30/35 (85%), Positives = 33/35 (93%)
40 Query: 1 MEADQVRGLLSADMLKHDIAMKKAVDVITSSATVK 35
 M ADQVR LLSADMLKHDIAMKKAV+VITS+A+VK
 Sbjct: 422 MPADQVRSLLSADMLKHDIAMKKAVEVITSTASVK 456

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2283

A DNA sequence (GBSx2417) was identified in *S.agalactiae* <SEQ ID 7039> which encodes the amino acid sequence <SEQ ID 7040>. This protein is predicted to be DNA-directed RNA polymerase, subunit delta. Analysis of this protein sequence reveals the following:

-2562-

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.2407(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:CAB15744 GB:Z99123 RNA polymerase (delta subunit) [Bacillus subtilis]
 Identities = 62/186 (33%), Positives = 102/186 (54%), Gaps = 15/186 (8%)
- Query: 1 MELEVFAGQEKSELSMIEVARAILEQRGRDNEMYFSDLVNDIQTYLGKSDSAIRESLPFF 60
 M ++ ++ +E E+++E+A + E+ + + F +L+N+I + LG + + + F
- 15 Sbjct: 1 MGIKQYSQEELKEMALVEIAHELFEHKKP--VPFQELLNEIASLLGVKKEELGDRIAQF 58
- Query: 61 YSDLNTDGSFIPLGENKWGLRSWYAIDEIDEIITLEEDEDGAPKRRKKRVNAFMDGDED 120
 Y+DLN DG F+ L + WGLRSWY D++DEE K KKK+ ++ D D
- 20 Sbjct: 59 YTDLNIDGRFLALSDQIWGLRSWYPYDQLDEE-----TQPTVKAKKKKAKKAVEEDLD 111
- Query: 121 AIDYNDDDPEDDFTEETPSLEYDEENPDDEKSEVESYDSEINEIIPDEDLDEDVEINEE 180
 ++ + D +D D E L+ + ++ D+E + + D EI E I DED DED
- 25 Sbjct: 112 LDEFBEIDEDDLDEVEEELDLEADDFDEEDLDEDDDDLEIEEDIIDED-DEDY----- 165
- Query: 181 DDEEEE 186
 DDEEEE
- Sbjct: 166 DDEEEE 171

- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7041> which encodes the amino acid
 sequence <SEQ ID 7042>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 35 bacterial cytoplasm --- Certainty=0.2263(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 40 Identities = 162/191 (84%), Positives = 181/191 (93%), Gaps = 1/191 (0%)
- Query: 1 MELEVFAGQEKSELSMIEVARAILEQRGRDNEMYFSDLVNDIQTYLGKSDSAIRESLPFF 60
 ++L+VFAGQEKSELSMIEVARAILE+RGRDNEMYFSDLVN+IQ YLGKSD+ IR +LPFF
- 45 Sbjct: 12 LKLDVFAGQEKSELSMIEVARAILEERGRDNEMYFSDLVNEIQNYLGKSDAGIRHALPFF 71
- Query: 61 YSDLNTDGSFIPLGENKWGLRSWYAIDEIDEIITLEEDEDGAPKRRKKRVNAFMDGDED 120
 Y+DLNTDGSFIPLGENKWGLRSWYAIDEIDEIITLEEDEDGA KRKKRVNAFMDGDED
- 50 Sbjct: 72 YTDLNTDGSFIPLGENKWGLRSWYAIDEIDEIITLEEDEDGAQKRRKKRVNAFMDGDED 131
- Query: 121 AIDYNDDDPEDDFTEETPSLEYDEENPDDEKSEVESYDSEINEIIPDEDLDEDVEINEE 180
 AIDY DDDPEDEDFTEE+ +EYDEE+PDDEKSEVESYDSE+NEIIP+D E+V+INEE
- 55 Sbjct: 132 AIDYRDDDPEDDFTEESA EYDEEDPDDEKSEVESYDSELNEIIPEDDF-EEVDINEE 190
- Query: 181 DDEEEEEEEEV 191
 D+E+EE+EE V
- Sbjct: 191 DEEDEEDEEPV 201

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2563-

Example 2284

A DNA sequence (GBSx2418) was identified in *S.galactiae* <SEQ ID 7043> which encodes the amino acid sequence <SEQ ID 7044>. This protein is predicted to be CTP synthetase (pyrG). Analysis of this protein sequence reveals the following:

```

5   Possible site: 23
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -0.11    Transmembrane    5 - 21 (    5 - 21)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAA09021 GB:AJ010153 CTP synthetase [Lactococcus lactis subsp.
      cremoris] (ver 2)
      Identities = 421/533 (78%), Positives = 481/533 (89%) .

20  Query: 2   TKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDYPINIDPGTMSPTYQHGEVYV 61
      TKYIFVTGG SS+GKGIVAASLGRLLKNRGLKVT+QKFDYP+NIDPGTMSPTYQHGEV+V
      Sbjct: 3   TKYIFVTGGTSSMGKGIVAASLGRLLKNRGLKVTVQKFDYPINIDPGTMSPTYQHGEV 62

      Query: 62   TDDGAETDLDLGHYERFIDINLNKYSNVTGKIYSEVLKERRGEYL GATVQVIPHVTDA 121
      TDDGAETDLDLGHYERFIDINLNKYSNVT+GK+YSE+L+KER+GEYL GATVQ++PHVT+
25  Sbjct: 63   TDDGAETDLDLGHYERFIDINLNKYSNVTSGKVYSEILRKERKGEYL GATVQMVPHVTNM 122

      Query: 122  LKEKIKRAATTTDSVDIITEVGGTVGDIESLPFLEALRQMKADVGSDNMVYIHTTLLPYL 181
      LKEKIKRAATTTD+D+IITEVGGTVGD+ESLPF+EALRQMK+VG+DNVMYIHT + +L
30  Sbjct: 123  LKEKIKRAATTTDADIITEVGGTVGDMESLPFLEALRQMKAEVGADNMVYIHTVPILHL 182

      Query: 182  KAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEQPAQOSIKNKLAQFCDAPEAVIESLD 241
      +AAGE+KTK Q++ K LR GIQ NMLV+R+E P +++K+A FCDVAPEAVI+SLD
      Sbjct: 183  RAAGELKTKIAQNATKTLREYGIQANMLVLRSEVPITTEMRDKIAMFCDAPEAVIQSLD 242

35  Query: 242  VDHIYQIPLNMQAQNMDQIVCDHLKLETPAADMTWSAMVDKVMNLEKKVKIALVGKYVE 301
      V+H+YQIPLN+QAQNMDQIVCDHLK+ P ADM EWSAMVD VMNL+KKVKIALVGKYVE
      Sbjct: 243  VEHLVQIPLNLQAQNMDQIVCDHLKLDAPKADMAEWSAMVDHVMNLKKVKIALVGKYVE 302

40  Query: 302  LPDAYLSVVEALKHSGYVNDVAIDLKWNAAEVTEDNIKELVGADAGIIVPGGFGQRGSE 361
      LPDAY+SV EALKH+GY +D +D+ WVNA +VT++N+ ELVGDA GIIVPGGFGQRG+E
      Sbjct: 303  LPDAYISVTEALKHAGYASDAEVDINWVNANDVTDENVAELVGDAAGIIVPGGFGQRGTE 362

      Query: 362  GKIEAIRYARENDVPMGLGVCLGMLTFCVEFARNVNLHGANSALDPKTPFPIDIMRDQ 421
      GKI AI+YARENDVPMGL+CLGMLT VEFARNVL L GA+S ELDP+T +P+IDIMRDQ
45  Sbjct: 363  GKIAAIKYARENDVPMGLGICLGMQLTAVEFARNVLGLEGAHSFELDPETKYPVIDIMRDQ 422

      Query: 422  IDIEDMGGLRLGLYPCKLKSGSRAAAAYNNQEVVQRRHRHRYEFNTKFREQFEAGFVF 481
      +D+EDMGGLRLGLYP KLK+GSRA AAYN+ EVVQRRHRHRYEFN K+RE FE AGFVF
50  Sbjct: 423  VDVEDMGGLRLGLYPAKLKNGSRAKAAAYNDAEVVQRRHRHRYEFNNKYREDFEKAGFVF 482

      Query: 482  SGVSPDNRLMEVVELPEKKFFVAAQYHPQLSRPNHAEELYTAFVTAAVENMK 534
      SGVSPDNRL+E+VEL KKFFVA QYHPQLSRPN EELYT F+ AVEN K
      Sbjct: 483  SGVSPDNRLVEIVELSGKKFFVACQYHPQLSRPNRPEELYTEFIRVAVENSK 535

```

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7045> which encodes the amino acid sequence <SEQ ID 7046>. Analysis of this protein sequence reveals the following:

```

      Possible site: 23
      >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -0.11    Transmembrane    5 - 21 (    5 - 21)

60  ----- Final Results -----
      bacterial membrane --- Certainty=0.1044(Affirmative) < succ>

```

-2564-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:CAA09021 GB:AJ010153 CTP synthetase [Lactococcus lactis subsp.
 cremoris] (ver 2)
 Identities = 423/532 (79%), Positives = 483/532 (90%)

10 Query: 2 TKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPIYNIDPGTMSPTYQHGEVYV 61
 TKYIFVTGG SS+GKGIVAASLGRLLKNRGLKVT+QKFDPIYNIDPGTMSPTYQHGEV+V
 Sbjct: 3 TKYIFVTGGGTSSMGKGIVAASLGRLLKNRGLKVTQKFDPIYNIDPGTMSPTYQHGEV 62

15 Query: 62 TDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKRRKGEYLGATVQVIPHITDA 121
 TDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKRRKGEYLGATVQ++PH+T+
 Sbjct: 63 TDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKRRKGEYLGATVQMVPHVTNM 122

20 Query: 122 LKEKIKRAASTTDSVDIITEVGGTVGDIESLPFLEALRQMKADVGSENVMIHTTLLPYL 181
 LKEKIKRAA+TTD+D+IITEVGGTVGD+ESLPF+EALRQMK+VG++NVMYIHT + +L
 Sbjct: 123 LKEKIKRAATTTDADIITEVGGTVGDMESLPFIEALRQMKAEVGADNVMYIHTVPILHL 182

25 Query: 182 KAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEEPVEQGIKNKLAQFCDVNSEAVIESRD 241
 +AAGE+KTK Q++ K LR GIQ NMLV+R+E P+ +++K+A FCDV EAVI+S D
 Sbjct: 183 RAAGELKTKIAQNATKTLREYGIQANMLVLRSEVPITTEMRDKIAMFCDVAPEAVIQSLD 242

30 Query: 242 VEHLQIPLNLQAQMDQIVCDHLKLNAPQADMTESAMVDKVMNLRKTKIALVGKYE 301
 VEHLQIPLNLQAQ+MDQIVCDHLKL+AP+ADM EWSAMVD VMNL+K KIALVGKYE
 Sbjct: 243 VEHLQIPLNLQAQNMMDQIVCDHLKLDAPKADMAEWSAMVDHVMNLKKVKIALVGKYE 302

35 Query: 302 LPDAYLSVVEALKHSGYANDTAIDLKVVNANDVTVDNAADLLGDADGIIVPGGFGQRGTE 361
 LPDAY+SV EALKH+GYA+D +D+ WVNANDVT +N A+L+GDA GIIVPGGFGQRGTE
 Sbjct: 303 LPDAYISVTEALKHAGYASDAEVDINWVNANDVTDENVAELVGDAAGIIVPGGFGQRGTE 362

40 Query: 362 GKIQAIRYARENDVFMGLGICLGMQLTCVEFARHVLNMEGANSFELEPSTKYPIIDIMRDQ 421
 GKI AI+YARENDVFMGLGICLGMQLT VEFAR+VL +EGA+SFEL+P TKYP+IDIMRDQ
 Sbjct: 363 GKIAAIKYARENDVFMGLGICLGMQLTAVEFARNVLGLEGAHSFELDPETKYPVIDIMRDQ 422

45 Query: 422 IDIEDMGGTLRLGLYPCKLKPGSKAAMAYNNQEVVQRRHRHRYEFNNKFRPEFEAGVFV 481
 +D+EDMGGTLRLGLYP KLK GS+A AYN+ EVVQRRHRHRYEFNNK+R +FE AGVFV
 Sbjct: 423 VDVEDMGGTLRLGLYPALKNGSRAKAAYNDAEVVQRRHRHRYEFNNKYREDFEKAGVFV 482

Query: 482 SGVSPDNRLVEIVELKEKKFFVAAQYHPELQSRPNRPEELYTAFVTAIKNS 533
 SGVSPDNRLVEIVEL KKKFFVA QYHPELQSRPNRPEELYT F+ A++NS
 Sbjct: 483 SGVSPDNRLVEIVELSGKKFFVACQYHPELQSRPNRPEELYTEFIRVAENS 534

An alignment of the GAS and GBS proteins is shown below.

Identities = 477/532 (89%), Positives = 503/532 (93%)

50 Query: 1 MTKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPIYNIDPGTMSPTYQHGEVY 60
 MTKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPIYNIDPGTMSPTYQHGEVY
 Sbjct: 1 MTKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPIYNIDPGTMSPTYQHGEVY 60

55 Query: 61 VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKRRKGEYLGATVQVIPHVT 120
 VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVL+KER+GEYLGATVQVIPH+TD
 Sbjct: 61 VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKRRKGEYLGATVQVIPHITD 120

60 Query: 121 ALKEKIKRAATTTDSVDIITEVGGTVGDIESLPFLEALRQMKADVGSNDVMYIHTTLLPY 180
 ALKEKIKRAA+TTDSVDIITEVGGTVGDIESLPFLEALRQMKADVGS+NVMYIHTTLLPY
 Sbjct: 121 ALKEKIKRAASTTDSVDIITEVGGTVGDIESLPFLEALRQMKADVGSENVMIHTTLLPY 180

65 Query: 181 LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEPQAGQSIKNKLAQFCDVAPEAVIESL 240
 LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTE+P Q IKNKLAQFCDV EAVIES
 Sbjct: 181 LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEEPVEQGIKNKLAQFCDVNSEAVIESR 240

Query: 241 DVDHIYQIPLNLQAQNMMDQIVCDHLKLETPAADMTESAMVDKVMNLEKKVKIALVGKYE 300
 DV+H+YQIPLN+QAQ+MDQIVCDHLKL P ADMTESAMVDKVMNL K KIALVGKYE

-2565-

Sbjct: 241 DVEHLYQIPLNLQAQSMQIVCDHLKLNAPQADMTSEMSAMVDKVMNLRKTTKIALVGKIV 300

Query: 301 ELPDAYLSVVEALKHSGYVNDVAIDLKWNAAEVTDNIKELVGADAGIIVPGGFGQRGS 360
ELPDAYLSVVEALKHSGY ND AIDLKWNAA +VT DN +L+GDADGIIVPGGFGQRG+

5 Sbjct: 301 ELPDAYLSVVEALKHSGYANDTAIDLKWNANDVTVDNAADLLGDADGIIVPGGFGQRGT 360

Query: 361 EGKIEAIRYARENDVPMGLVCLGMQLTCVEFARNVLNLHGANSALDPKTPFPPIIDIMRD 420
EGKI+AIRYARENDVPMGL+CLGMQLTCVEFAR+VLN+ GANS EL+P T +PIIDIMRD

10 Sbjct: 361 EGKIQAIRYARENDVPMGLGICLGMQLTCVEFARHVLNMEGANSFELEPSTKYPIIDIMRD 420

Query: 421 QIDIEDMGGTLRLGLYPCKLKSGSRAAAAYNNQEVVQRRHRHRYEFNTKFREQFEAGFV 480
QIDIEDMGGTLRLGLYPCKLK GS+AA AYNNQEVVQRRHRHRYEFN KFR +FEAGFV

Sbjct: 421 QIDIEDMGGTLRLGLYPCKLKPGSKAAMAYNNQEVVQRRHRHRYEFNNKFRPEFEAGFV 480

15 Query: 481 FSGVSPDNRLMEVVELPEKKFFVAAQYHPELQSRPNHAEELYTAFVTAIVEN 532
FSGVSPDNRL+E+VEL EKKFFVAAQYHPELQSRPN EELYTAFVTA++N

Sbjct: 481 FSGVSPDNRLVEIVELKEKKFFVAAQYHPELQSRPNRPEELYTAFVTAIAKN 532

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2285

A DNA sequence (GBSx2419) was identified in *S.agalactiae* <SEQ ID 7047> which encodes the amino acid sequence <SEQ ID 7048>. Analysis of this protein sequence reveals the following:

Possible site: 34

25 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -9.92 Transmembrane 13 - 29 (3 - 34)

----- Final Results -----

30 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9285> which encodes amino acid sequence <SEQ ID 9286> was also identified.

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14296 GB:Z99116 yqkD [Bacillus subtilis]
Identities = 79/289 (27%), Positives = 139/289 (47%), Gaps = 8/289 (2%)

40 Query: 1 MKKIRLSKFIKMIVVILFLISVAASFYFFHVAQVRDDKSFISNGQRKPGNSLYAYDKSFD 60
MKKI L+ I +V + I + S ++ D+ I + G+ ++ +SF+

Sbjct: 1 MKKILLA--IGALVTAVIAIGIVFSHMLFIKKTDED--IIKRETDNGHDVF--ESFE 53

Query: 61 KLLKQKIEMTNQNIKQVAWYVPAVKKTHKTAVVHGFANSKENMKAYGWLPHKLGYNVLM 120
++ K + + + Y A T T ++ HG + N Y LF LG+NVL+

45 Sbjct: 54 QMEKTAFFVIPSAYGYDIKGYHVAPHDTPTNTIIICHGVTMNVLSLKMHLFLDLGWNVLI 113

Query: 121 PDNIAHGESHQGLIGYGWNDRENI IKWTEMIVDK- NPSSQITLFGVSMGGATVMMASGEK 179
D+ HG+S G+ YG+ +++++ K ++ +K N I + G SMG T ++ +G

50 Sbjct: 114 YDHRRHGQSGGKTSYGYEKDDLKVVSLKKNKTNRHGLIGIHGESMGAVTALLYAGAH 173

Query: 180 LPSQVNNIIEDCGYSSVWDELKFOAKEMYGLPAFPLLYEVSTISKIRAGFSYQGQASSVEQ 239
I DC ++ ++L ++ + Y LP++PLL K+R G+ + S +

Sbjct: 174 CSDGADFYIADCPFCFDEQLAYRLRAEYRLPSWPLLPIADFFLKLKRGYRAREVSPPLAV 233

55 Query: 240 LKKNNLPALFIHGDKDNFVPTSMVYDNYKATAGKKELYIVKGAKHAKSF 288
+ K P LFIH D+++P S Y+ G K LYI + +HA S+

Sbjct: 234 IDKIEKPVLFHISKDDDYIPVSSTERLYEKRGPKALYIAENGEHMSY 282

-2566-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7049> which encodes the amino acid sequence <SEQ ID 7050>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq
 5 INTEGRAL Likelihood = -7.48 Transmembrane 10 - 26 (3 - 32)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3994(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 10 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14296 GB:Z99116 yqkD [Bacillus subtilis]
 Identities = 88/295 (29%), Positives = 145/295 (48%), Gaps = 4/295 (1%)
 15 Query: 10 LGILFLITLISVGASFYFFHVAQIREEEKSFNNKKRSTNNPLYPAEQSFDAIPYEKRL 69
 L I L+ +I++G F.H+ ++K+ + KR T+N + +SF+ + +
 Sbjct: 6 LAIGALVTAVIAIG--IVFSHMLIFIKKKTDEDIIKRETNG-HDVFESFEQMEKTAFTVI 62
 20 Query: 70 TNRGLKQVGWYLPAAQTKTKTAIVVHGFTNDKEDMKPYAMLFHDLGYNVIMPDNEAHGES 129
 + + Y A T T I+ HG T + + Y LF DLG+NVL+ D+ HG+S
 Sbjct: 63 PSAYGYDIKGYHVAPHDTPTNTIIICHGVTMNVLSLKYMHFLDLGWNVLIYDHRRHGQS 122
 25 Query: 130 EGNLIGYGNDRNLNVMAWTDQLI-KENPESQITLFGLSMGATVMMASGERLPAQVTSLI 188
 G YG+ ++ ++ L K N I + G SMGA T ++ +G I
 Sbjct: 123 GGGTTSYGYEYKDDLNVSVLLKNTNHRGLIGHGESMGAVTALLYAGAHCSGDADFYI 182
 Query: 189 EDCGYASVWDELKFKQAKAMYNLPAFPLLYEVSALSIRAGFSYGEASSVKQLAKNKRPTL 248
 DC +A ++L ++ +A Y LP++PLL K+R G+ E S + + K ++P L
 30 Sbjct: 183 ADCPFACFDEQLAYRLRAEYRLPSWPLLPIADFFLKLKRGYRAREVSPLAVIDKIEKPEVL 242
 Query: 249 FIHGDKDDFVPTKMVDNYKATKGPKKILIVKGAKHAKSFETNPEQYQKKIAAFL 303
 FIH DD++P Y+ +GPK + I + +HA S+ N Y+K + FL
 Sbjct: 243 FIHSDDDYIPVSSTERLYEKKRGPKALYIAENGHAMSYTKNRHTYRKTQVEFL 297
 35

An alignment of the GAS and GBS proteins is shown below.

Identities = 203/294 (69%), Positives = 246/294 (83%)
 40 Query: 1 MKKIRLSKFIKMIVILFLISVAASFYFFHVAQVRDDKSFIENGQRKPGNSLYAYDKSFD 60
 MK IR++K++ ++ +++ LISV ASFYFFHVAQ+R++KSFI+N +R N LY ++SFD
 Sbjct: 1 MKTIRIAKYLIGILFLITLISVGASFYFFHVAQIREEEKSFNNKKRSTNNPLYPAEQSFDA 60
 Query: 61 KLLKQKIEMTNQNIQVAVYVPAVKKTHKTAVVHGFANSKENMKAYGWLPHKLGYNVLM 120
 L +K ++TN+ +KQV WY+PA +KT KTA+VHGF N KE+MK Y LFH LGYNVLM
 45 Sbjct: 61 ALPYEKRLTNRGLKQVGWYLPAAQTKTKTAIVVHGFTNDKEDMKPYAMLFHDLGYNVLM 120
 Query: 121 PDNIAHGESHGQLIGYGNDRNIIKWTEMIVDKNPSSQITLFGVSMGGATVMMASGEKL 180
 PDN AHGES G LIGYGNDR N++ WT+ ++ +NP SQITLFG+SMG ATVMMASGE+L
 50 Sbjct: 121 PDNEAHGESEGNLIGYGNDRNLNVMAWTDQLIKENPESQITLFGLSMGATVMMASGERL 180
 Query: 181 PSQVNNIIEGCGYSSVWDELKFKQAKEMYGLPAFPLLYEVSTISKIRAGFSYQASSVEQL 240
 P+QV ++IEDCGY+SVWDELKFKQAK MY LPAPPLLYEVS +SKIRAGFSYG+ASSV+QL
 Sbjct: 181 PAQVTSIIEGCGYASVWDELKFKQAKAMYNLPAFPLLYEVSALSIRAGFSYGEASSVKQL 240
 55 Query: 241 KKNLPAIFIHGDKDNFVPTSMVDNYKATAGKKELYIVKGAKHAKSFETEPEK 294
 KN P LFIHGDKD+FVPT MVDNYKAT G KE+ IVKGAKHAKSFET PE+
 Sbjct: 241 AKNKRPTLFIHGDKDDFVPTKMVDNYKATKGPKKILIVKGAKHAKSFETNPEQ 294

60 SEQ ID 9286 (GBS662) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 8-10; MW 63kDa) and in Figure 187 (lane 4; MW 63kDa).

GBS662-GST was purified as shown in Figure 237, lane 7.

-2567-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2286

A DNA sequence (GBSx2420) was identified in *Sagalactiae* <SEQ ID 7051> which encodes the amino acid sequence <SEQ ID 7052>. This protein is predicted to be aspartate--ammonia ligase (asnA). Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2898(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9309> which encodes amino acid sequence <SEQ ID 9310> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
Identities = 246/300 (82%), Positives = 268/300 (89%)

Query: 1 MIDKLEIVEVQGPILSQVGDGMQDNLSGIEHPVSVKVINIPEAEFEVVHSLAKWKRHTLA 60
+I++L I+EVQGPILSQVG+GMQDNLSGIE V V V IP A FEVVHSLAKWKRHTLA
Sbjct: 23 LIEQLGIIEVQGPILSQVGNQMQLNSGIEKAVQVNVKCI PNAVFEVVHSLAKWKRHTLA 82

Query: 61 RFGFNEGEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPDGRNLDYLKETVEKIYK 120
RF F E EGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIP+GRRN YLKETV IY+
Sbjct: 83 RFNFKEDEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPEGRRNFAYLKETVNSIYR 142

Query: 121 AIRLTELAVEARFDIESILPKRITFIHTEELVEKYPDLSPKERENAIKEYGAVFLIGIG 180
AIRLTELAVEARFDI SILPK+ITF+H+E+LV++YDLS KERENAI KEYGAVFLIGIG
Sbjct: 143 AIRLTELAVEARFDIPSILPKQITFVHSEDLVKRYPDLSKKERENAIKEYGAVFLIGIG 202

Query: 181 GELADGKPHDGRAPDYDDWTTTSENGFKLNGDILVWNEQLGTAFELSSMGIRVDEDAK 240
G+L+DGKPHDGRAPDYDDWTT SENG+KGLNGDILVWN+QLG AFELSSMGIRVDE AL+
Sbjct: 203 GKLSGDKPHDGRAPDYDDWTTTSENGYKGLNGDILVWNDQLGKAFELSSMGIRVDESALR 262

Query: 241 RQVVLTGDEDRLEFEWHKTLRLRGFFPLTIGGGIGQSRLAMFLLRKXHIGEVQSSVWPKEV 300
QV LTGDED L+ +WH+ LL G PLTIGGGIGQSRLAM LLRK HIGEVQSSVWPKE+
Sbjct: 263 LQVGLTGDEHKLKMDWHQDLLNGKLPLTIGGGIGQSRLAMLLLRKKHIGEVQSSVWPKEV 322

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7053> which encodes the amino acid sequence <SEQ ID 7054>. Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 189 - 205 (189 - 205)

----- Final Results -----
bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
Identities = 255/330 (77%), Positives = 289/330 (87%)

Query: 1 MKKSFHQEEISFVKNTFTQYLIKLDVVEVQGPILSRVGDGMQDNLSGTENPVSVNVL 60
MKK+FI QQ+EISFVKNTFTQ LI +L ++EVQGPILS+VG+GMQDNLSG E V VNV

-2568-

Sbjct: 1 MKKTFFILQQQEISFVKNFTQNLIEQLGIIEVQGPILSQVGNMGMDNLSGIEKAVQVNVK 60

Query: 61 KIPNATFEVVHSLAKWKRHTLARFGFNEGEGLVNMKALRPDEDSLDQTHSVYVDQWDWE 120
IPNA FEVVHSLAKWKRHTLARF F E EGL V+MKALRPDEDSLD THSVYVDQWDWE

5 Sbjct: 61 CIPNAVFEVVHSLAKWKRHTLARFNFKEDEGLFVHMKALRPDEDSLDPTHSVYVDQWDWE 120

Query: 121 KVIPDGKRNLAYLKETVETIYKVIIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDL 180
KVIP+G+RN AYLKETV +IY+ IRLTELAVEAR+DI ++LPK+ITF+H+E+LV +YDDL

10 Sbjct: 121 KVIPEGRRNFAYLKETVNSIYRAIRLTELAVEARFDIPSILPKQITFVHSEDLVKRYDDL 180

Query: 181 TPKERENAITKEFGAVFLIGIGGVLDPGKPHDGRAPDYDDWTETENG YHGLNGDILVWN 240
+ KERENAI KE+GAVFLIGIGG L DGKPHDGRAPDYDDWTTE+ENGY GLNGDILVWN

Sbjct: 181 SSKERENAIKEYGAVFLIGIGGKLSDGKPHDGRAPDYDDWTTESENGYKGLNGDILVWN 240

15 Query: 241 DQLGSFAFELSSMGIRVDEEALRQVEMTGDQDRLGFDWHKSLNLGLFPLTIGGGIGQSRM 300
DQLG AFELSSMGIRVDE AL+ QV +TGD+D L DWH+ LLNG PLTIGGGIGQSR+

Sbjct: 241 DQLGAFELSSMGIRVDESALRLQVGLTGDDEHLKMDWHQDLLNGKLPLTIGGGIGQSRL 300

Query: 301 VMFLLRKQHIGEVQTSVWPQEVDRDSYDNL 330
M LLRK+HIGEVQ+SVWP+E+ + + NIL

20 Sbjct: 301 AMLLLRKKHIGEVQSSVWPKEMLEEFSNL 330

An alignment of the GAS and GBS proteins is shown below.

Identities = 254/303 (83%), Positives = 280/303 (91%)

25 Query: 1 MIDKLEIVEVQGPILSQVGDGMQDNLSGIEHPVSVKVLNIPAEFEVVHSLAKWKRHTLA 60
+I'KL++VEVQGPILS+VGDGMQDNLSG E+PVS V L IP A FEVVHSLAKWKRHTLA

Sbjct: 23 LIAKLDVVEVQGPILSRVGDGMQDNLSGTENPVSNNVLKIPNATFEVVHSLAKWKRHTLA 82

30 Query: 61 RFGFNEGEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPDGRRNLDYLKETVEKIYK 120
RFGFNEGEGL V+MKALRPDEDSLD THSVYVDQWDWEKVIPDG+RNL YLKETVE IYK

Sbjct: 83 RFGFNEGEGLVNMKALRPDEDSLDQTHSVYVDQWDWEKVIPDGKRNLAYLKETVETIYK 142

35 Query: 121 AIRLTELAVEARFDIESILPKRITFIHTEELVEKYPDLSPKERENAIKEYGAVFLIGIG 180
IRLTELAVEAR+DIE++LPK+ITFIHTEELV KYPDL+PKERENAI KE+GAVFLIGIG

Sbjct: 143 VIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDLTPKERENAITKEFGAVFLIGIG 202

Query: 181 GELADGKPHDGRAPDYDDWTTPSENGFKGLNGDILVWNEQLGTAFELSSMGIRVDEEALK 240
G L DGKPHDGRAPDYDDWT +ENG+ GLNGDILVWN+QLG+AFELSSMGIRVDE+ALK

40 Sbjct: 203 GVLDPGKPHDGRAPDYDDWTETENG YHGLNGDILVWNDQLGSFAFELSSMGIRVDEEALK 262

Query: 241 RQVVLTGDEDRLEFEWHKTLRLRGFFPLTIGGGIGQSRLAMFLLRKXHIGEVQSSVWPKEV 300
RQV +TGD+DRL F+WHK+LL G FPLTIGGGIGQSR+ MFLLRK HIGEVQ+SVWP+EV

45 Sbjct: 263 RQVEMTGDQDRLGFDWHKSLNLGLFPLTIGGGIGQSRMVMFLLRKQHIGEVQTSVWPQEV 322

Query: 301 RDT 303
RD+

Sbjct: 323 RDS 325

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2287

A DNA sequence (GBSx2421) was identified in *S.agalactiae* <SEQ ID 7055> which encodes the amino acid sequence <SEQ ID 7056>. Analysis of this protein sequence reveals the following:

55 Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.3163(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2569-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2288

A DNA sequence (GBSx2422) was identified in *S.galactiae* <SEQ ID 7057> which encodes the amino acid sequence <SEQ ID 7058>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

10

```
----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

15

A related GBS nucleic acid sequence <SEQ ID 9007> which encodes amino acid sequence <SEQ ID 9008> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
Identities = 30/97 (30%), Positives = 50/97 (50%), Gaps = 3/97 (3%)

Query: 50 KALVSKSQQSEATIFIGRPTCQYCRFLPKLLKSQATLHSKIYYLDSQKYKG-KRLKSFF 108
      +A + ++ AT FIGR TC YCR F L A + IY+++S++ L++F
Sbjct: 18 RAQEALDKKETATFFIGRKTCPCYCRKFAGTSLGTVVAETKAHIYFINSEEASQLNDLQAFR 77

Query: 109 KKHHTITVPNLAHYQQGKMTKYLQGSQATPQQIQTF 145
      ++ I TVP H G++ + S + Q+I+ F
Sbjct: 78 SRYGIPTVPGFVHITDGQIN--VRCDSSMSAQEIKDF 112
```

30 SEQ ID 9008 (GBS134) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 2; MW 17kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 4; MW 42kDa).

GBS134-GST was purified as shown in Figure 204, lane 10.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2289

A DNA sequence (GBSx2423) was identified in *S.galactiae* <SEQ ID 7059> which encodes the amino acid sequence <SEQ ID 7060>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0735(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

45

A related GBS nucleic acid sequence <SEQ ID 9603> which encodes amino acid sequence <SEQ ID 9604> was also identified.

-2570-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06309 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 78/178 (43%), Positives = 115/178 (63%), Gaps = 3/178 (1%)

5 Query: 3 MRVVAGTFGGRPLKTLTGKTRPTTDKVKGAIFNMIGPFFEGGRVLDLFSGSGSLAIEAI 62
MRV+AG G LK + G TRPTTDKVK AIFNMIGPFF+GG LDL+ GSG L IEA+
Sbjct: 1 MRVIAGEQKGLTLKAVPGHKTRPTTDKVKBAIFNMIGPFFDGGIGLDLYGSGGGLGIEAL 60

10 Query: 63 SRGMDQAVLVEKDRRAQVVIQENIAMTKSPEQFQLLKMEANRALEQLTGQ---FDLVLLD 119
SRG+++ + V++ +RA I++N++ + ++ + +A RAL+ LT + F V LD
Sbjct: 61 SRGVERMIFVDQOKRAIETIKQNLSCGLEGRAEVYRNDAKRALQVLTKRGIVFAYVFLD 120

Query: 120 PPYAKEEIVKQIQIMDSKGLLGDIMACETDKSVLDPEEIASFGIWKQKIYGISKVT 177
PPYAK+ I + I+ + GLL + ++ CE D+ LP++I K++ YG + +T
15 Sbjct: 121 PPYAKQTIKNDLAILANHGLLEEGGVVCEHDRDTMLPDQIEYAVKHKEETYGDTMIT 178

There is also homology to SEQ ID 132.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2290

A DNA sequence (GBSx2424) was identified in *S.agalactiae* <SEQ ID 7061> which encodes the amino acid sequence <SEQ ID 7062>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4984(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB96619 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
Identities = 175/254 (68%), Positives = 219/254 (85%)

35 Query: 2 LRRHIYSMLLEHXHLQPEIKYHQKTNLKRNRYTVFIEEKVDVILADLKLADAFPGIETG 61
L RH+Y ++ EI++HQ++NLRKNRYTVF +EKV +L+DL LAD+FFG+ETG
Sbjct: 50 LARHLYESFLHFYEIKSEIRHHQRSNLRKNRYTVFTDEKVDLLSDLHLADSFFGLETG 109

40 Query: 62 IEHSILDNDENGRAYLRGAFLSTGTVREPDGSKYQLEIFSVYLDHAQDLANLMMKFMFLDA 121
I+ +IL ++E GRAYL GAFL+ G++R+P+SGKYQLEI SVYLDHAQ +A+L+++F+LDA
Sbjct: 110 IDEAILSDEEAGRAYLCGAFLANGSIRDPEGSKYQLEISSVYLDHAQGIASLLQQFLDLA 169

Query: 122 KVIEHKHGAVTYLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIA 181
KV+E K GAVTYLQ+AEDIMDFLIVI AM+ARD FE +K++RETRND+NRANN ETANIA
45 Sbjct: 170 KVLERKKGAVTYLQRAEDIMDFLIVIGAMQARDDEFERVKILRETRNDLNRANNAETANIA 229

Query: 182 RTITASMKTTINNIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGV 241
RT++ASMKTINNI KI D +G + LP DL++VAQ+R+ HPDYSIQQ+ADSL TPL+KSGV
Sbjct: 230 RTVSASMKTINNISKIKDIMGLENLPVDLQEVQRLRIQHPDYSIQQIADSLETPLTKSGV 289

50 Query: 242 NHRLRKINKIADEL 255
NHRLRKINKIADEL
Sbjct: 290 NHRLRKINKIADEL 303

55 There is also homology to SEQ ID 5540:

Identities = 186/254 (73%), Positives = 227/254 (89%)

Query: 2 LRRHIYSMLLEHXHLQPEIKYHQKTNLKRNRYTVFIEEKVDVILADLKLADAFPGIETG 61
+ R+IYS++E+ + PEI+YHQKTNLKRNRYTV++E+ V+ ILADLKLAD+FFG+ETG

-2571-

Sbjct: 50 IARYIYSLIEDAYVIVPEIRYHQKTNLKRNRYTVYVEQGVETILADLKLADSFFGLETG 109

Query: 62 IEHSILDNDENGRAYLRGAFLSTGTVREPDGKYQLEIFSVYLDHAQDLANLMKKFMLDA 121
IE +L +D GR+YL+GAFL+ G++R+P+SGKYQLEI+SVYLDHAQDLA LM+KFMLDA

5 Sbjct: 110 IEPQVLSDDNAGRSYLGGAFLAAGSTRDPESGKYQLEIYSVYLDHAQDLAQLMQKFMLDA 169

Query: 122 KVIEHKHGAVTYLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIA 181
K IEHK GAVTYLQKAEDIMDFLI+I AM ++ FE IK++RE RNDINRANN ETANIA

10 Sbjct: 170 KTIEHKSGAVTYLQKAEDIMDFLIIGAMSCKEDEFAIKLLREARNNDINRANNAETANIA 229

Query: 182 RTITASMTINNIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGV 241
+TI+ASMTINNIIKIMDTIG ++LP +L+QVAQ+RV HPDYSIQQ+AD+LE P++KSGV

15 Sbjct: 230 KTISASMTINNIIKIMDTIGLESLEPIELQQVAQLRVKHPDYSIQQVADALEFPITKSGV 289

Query: 242 NHRLRKINKIADEL 255
NHRLRKINKIAD+L

Sbjct: 290 NHRLRKINKIADDL 303

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 2291

A DNA sequence (GBSx2425) was identified in *S.agalactiae* <SEQ ID 7063> which encodes the amino acid sequence <SEQ ID 7064>. Analysis of this protein sequence reveals the following:

Possible site: 14
25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0297(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
35 vaccines or diagnostics.

Example 2292

A DNA sequence (GBSx2428) was identified in *S.agalactiae* <SEQ ID 7065> which encodes the amino acid sequence <SEQ ID 7066>. Analysis of this protein sequence reveals the following:

Possible site: 31
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB54571 GB:AJ006393 response regulator [Streptococcus pneumoniae]
Identities = 139/190 (73%), Positives = 166/190 (87%)

50 Query: 8 IKIVLVDDHEMVRGLKSFNLQADVEVIGEASNGLEGIKKALELRPDVVMDLVMPEND 67
+KI+LVDDHEMVRGLKSL+ +LQ DVEV+GEASNG +GI ALELRPDV+VMD+VMPEN+
Sbjct: 1 MKILLVDDHEMVRGLKSYFDLQDDVEVVGESNGSQGIDLALALELRPDVIVMDIVMPEN 60

-2572-

Query: 68 GVEATLALLKDWPEAAILVLTSLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKVSRG 127
 G++ATLA+LK+WPEA IL++TSYLDNEKI PV++AGAKGYMLKTSSA E+L+A+ KV+ G
 Sbjct: 61 GIDATLAILKEWPEAKILIVTSLDNEKIMPVLDAGAKGYMLKTSSADELLHAVSKVAAG 120

5 Query: 128 EQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTIVKTHV 187
 E AIE EV KK++ H LHE LTARERD+L L+AKGY+NQRIAD+LFISLKTIVKTHV
 Sbjct: 121 ELAIEQEVSKKVEYHRNHMELHEELTARERDVLQLIAGYENQRIADDLFISLKTIVKTHV 180

10 Query: 188 SNILGKLNGS 197
 SNIL KL S
 Sbjct: 181 SNILAKLEVS 190

There is also high homology to SEQ ID 2996:

Identities = 158/198 (79%), Positives = 176/198 (88%), Gaps = 1/198 (0%)

15 Query: 5 MDKIKIVLVDDHEMVRGLKSFNLQADVEVIGEASNGLEGIKKALELRPDVVMDLVMP 64
 M KIK++LVDDHEMVR+GLKSFNLQAD++V+GEASNG EG+ AL L+PDV+VMDLVMP
 Sbjct: 3 MSKIKVILVDDHEMVRMGLKSFNLQADIDVVGEASNGREGVDLALALKPDVLVMDLVMP 62

20 Query: 65 EMDGVEATLALLKDWPEAAILVLTSLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKV 124
 E+ GVEATL +LK W EA +LVLTSLDNEKIYPVI+AGAKGYMLKTSSAAEILNAIRKV
 Sbjct: 63 ELGGVEATLEVLLKKWKEAKVLVLTSLDNEKIYPVIDAGAKGYMLKTSSAAEILNAIRKV 122

25 Query: 125 SRGEQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTIVK 184
 S+GE AIE EVDKKIKAHD+ P LHE LTARE DIL+LLAKGYDNQ IADELFISLKTIVK
 Sbjct: 123 SKGELAIEETEVDKKIKAHDQHPDLHEELTAREVDILHLLAKGYDNQTIADLFISLKTIVK 182

Query: 185 THVSNILGKLN-GSRSNS 201
 THVSNIL KL G R+ +
 30 Sbjct: 183 THVSNILAKLEVGDRTQA 200

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2293

35 A DNA sequence (GBSx2429) was identified in *Sagalactiae* <SEQ ID 7067> which encodes the amino acid sequence <SEQ ID 7068>. This protein is predicted to be histidine kinase (narQ). Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3944(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB54570 GB:AJ006393 histidine kinase [Streptococcus pneumoniae]
 Identities = 32/55 (58%), Positives = 49/55 (88%)

50 Query: 1 MIDNGIGFDMDSVYDLSYGLKNIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQ 55
 ++DNGIGF + S+ DLSYGL+NI++RVED+AG +QLL+ P +G+A+DIR+PL+++
 Sbjct: 276 VVDNGIGFQLGSLDDLSYGLRNIKERVEDMAGTVQLLTAPKQGLAVDIRIPLLDK 330

There is also homology to SEQ ID 2992:

55 Identities = 44/59 (74%), Positives = 51/59 (85%)

Query: 1 MIDNGIGFDMDSVYDLSYGLKNIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQSEDK 59
 MID+G+GFDM V DLSYGLKNIEDRV DLAGNL L+SQ GKGV+MDIRLP+V +D+
 Sbjct: 276 MIDDGVGFDMQVRDLSYGLKNIEDRVNDLAGNLHLISQKGVSMDIRLPIVKGDDDE 334

-2573-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2294

- 5 A DNA sequence (GBSx2430) was identified in *S.agalactiae* <SEQ ID 7069> which encodes the amino acid sequence <SEQ ID 7070>. This protein is predicted to be RfbQRSO155-1. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

10

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1120(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

15

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 7072:

Identities = 171/172 (99%), Positives = 172/172 (99%)

20

Query: 1 MGQVAVEEKSNEIVAIPQLLRITIDIRKSIVTIDAMGTQTAIVDTTIKKGADYCLAVKGNQ 60

+GQVAVEEKSNEIVAIPQLLRITIDIRKSIVTIDAMGTQTAIVDTTIKKGADYCLAVKGNQ

Sbjct: 143 LGQVAVEEKSNEIVAIPQLLRITIDIRKSIVTIDAMGTQTAIVDTTIKKGADYCLAVKGNQ 202

25

Query: 61 ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYVWSSDIKWLCQNHKPKWHK 120

ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYVWSSDIKWLCQNHKPKWHK

Sbjct: 203 ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYVWSSDIKWLCQNHKPKWHK 262

30

Query: 121 LRIGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFCNCVRGHWQIESMHWLL 172

LRIGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFCNCVRGHWQIESMHWLL

Sbjct: 263 LRIGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFCNCVRGHWQIESMHWLL 314

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2295

- 35 A DNA sequence (GBSx2431) was identified in *S.agalactiae* <SEQ ID 7073> which encodes the amino acid sequence <SEQ ID 7074>. This protein is predicted to be translation initiation factor if-3 homolog dsq (infC). Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

40

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1787(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA68920 GB:Y07640 translation initiation factor, IF3 [Listeria monocytogenes]

Identities = 112/169 (66%), Positives = 134/169 (79%)

50

Query: 7 KDLFINDEIRVREVRVLVGLGEQGLGIKPLSEAQAIAADDANVDLVLIQPOATPPVAKIMDY 66

KD+ +ND IR REVRL+ +GEQLG+K +A IA+ AN+DLVL+ P A PPVA+IMDY

Sbjct: 3 KDMLVNDGIRAREVRVLIDQDGEQGLGVKSKIDALQIAKANLDDLVLVAPTAKPPVARIMDY 62

Query: 67 GKFKFEYQKKQKEQRKKQSVTVKEVRVSPVIDKGDFFETKLRNGRKFLKGNKVKVSIRF 126

GKF+FE QKK KE RK Q V+ +KEVRVSP ID+ DF+TKLRN RKFLKKG+KVK SIRF

-2574-

Sbjct: 63 GKFRFEQQKKDKKARKNQKVIVMKEVRLSPTIDEHDFDTKLRNARKFLEKGDVKCSIRF 122

Query: 127 KGRMITHKEIGAKVLAFAEATQDIAIEQRAKMDGRQMFQMLAPIPDK 175
KGR ITHKEIG KVL FA+A +D+ IEQR KMDGR MF+ LAP+ +K

Sbjct: 123 KGRAITHKEIGQKVLDRFAKACEDLCTIEQRPKMDGRSMFLVLAPLHEK 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7075> which encodes the amino acid sequence <SEQ ID 7076>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2247(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 167/176 (94%), Positives = 173/176 (97%)

Query: 1 MKIIAKKDLFINDEIRVREVLVGLGEQQLGIKPLSEAQAIAADDANVDLVLIQPQATPFV 60

+KIIAKKDLFINDEIRVREVLVGLGEQQLGIKPLSEAQ++AD +NVDLVLIQPQA PFV

Sbjct: 1 VKIIAKKDLFINDEIRVREVLVGLGEQQLGIKPLSEAQSLADASNVDLVLIQPQAVPFV 60

Query: 61 AKIMDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGFETKLRNGRKFFLEKGNKV 120

AK+MDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGFETKLRNGRKFFLEKGNKV

Sbjct: 61 AKIMDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGFETKLRNGRKFFLEKGNKV 120

Query: 121 KVSIRFKGRMITHKEIGAKVLAFAEATQDIAIEQRAKMDGRQMFQMLAPIPDKK 176

KVSIRFKGRMITHKEIGAKVLA+FAEATQDIAIEQRAKMDGRQMFQMLAPI DKK

Sbjct: 121 KVSIRFKGRMITHKEIGAKVLADFAEATQDIAIEQRAKMDGRQMFQMLAPISDKK 176

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2296

A DNA sequence (GBSx2432) was identified in *S.agalactiae* <SEQ ID 7077> which encodes the amino acid sequence <SEQ ID 7078>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1807(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45308 GB:U81957 RNA polymerase beta' subunit [Streptococcus gordonii]
Identities = 262/286 (91%), Positives = 276/286 (95%)

Query: 1 MAAKVVKAGVBEVXIRSVFTCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG 60

MA +VV AGV EV IRSV TCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG

Sbjct: 122 MARQVFNAGVTEVTIRSVLTNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG 181

Query: 61 TQLTMRFTFHTGGVASNTDITQGLPRIQEIFEARNPKGEAVITEVKGEVVAIEEDSSTRTK 120

TQLTMRFTFHTGGVAS++DITQGLPR+QEIFEARNPKGEAVITEVKGEV AIEED+STRTK

Sbjct: 182 TQLTMRFTFHTGGVASSSDITQGLPRVQEIFEARNPKGEAVITEVKGEVTAIEEDASTRTK 241

Query: 121 KVFVKGQTGEGEYVVPFTARMKVEVGDEVARGAALTEGSIQPKRLLEVRDTLSVETYLLA 180

KVFVKGQTGEGEYVVPFTARMKVEVG+V+RGAALTEGSIQPK LL VRD LSVETYLLA

Sbjct: 242 KVFVKGQTGEGEYVVPFTARMKVEVG+V+RGAALTEGSIQPKHLLAVRDLVSVETYLLA 301

-2575-

Query: 181 EVQKVYRSQGVEIGDKHVEVMVRQMLRKVRVMDPGDSDLPGTLMDISDFTDANKDIVIS 240
 EVQKVYRSQGVEIGDKH+EVMVRQM+RKVRVMDPGDSDL GTLMDI+DFTDAN+D+VIS
 Sbjct: 302 EVQKVYRSQGVEIGDKHIEVMVRQMLRKVRVMDPGDSDLPGTLMDISDFTDANRDVVIS 361

Query: 241 GGIPATSRPVLMSGITKASLETNSFLSAASFQETTRVLTDAAIRGKK 286
 GG+PAT+RPVLMSGITKASLETNSFLSAASFQETTRVLTDAAIRGKK
 Sbjct: 362 GGVPTARPVLMSGITKASLETNSFLSAASFQETTRVLTDAAIRGKK 407

10 There is also homology to SEQ ID 384.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2297

15 A DNA sequence (GBSx2434) was identified in *S.agalactiae* <SEQ ID 7079> which encodes the amino acid sequence <SEQ ID 7080>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0352(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2298

30 A DNA sequence (GBSx2435) was identified in *S.agalactiae* <SEQ ID 7081> which encodes the amino acid sequence <SEQ ID 7082>. This protein is predicted to be acetoin dehydrogenase (TPP-dependent) beta chain (pdhB). Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0266(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAE04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
 chain [Bacillus halodurans]
 Identities = 37/57 (64%), Positives = 50/57 (86%)

45 Query: 1 MLEEFQAKRVRDTPISEAAIAGSAIGAAQTGLRPIVDLTFMDFVTIAMDAIVDDCIR 57
 M+EEFG++RVR+TPISEAAI+G+AIGAA TG+RPI++L F DF+TIAMD +V+ +
 Sbjct: 44 MLEEFQSERVRNTPISEAAISGTAIGAAITGMREILELQFSDFTTIAMDNMVNQAQK 100

There is also homology to SEQ ID 4272.

-2576-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2299

A DNA sequence (GBSx2436) was identified in *S.agalactiae* <SEQ ID 7083> which encodes the amino acid sequence <SEQ ID 7084>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3015(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB18706 GB:U38906 Structural protein [Bacteriophage rlt]
Identities = 57/127 (44%), Positives = 83/127 (64%)

Query: 5 IKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEGEQKL 64
+ GTLF P LVT+++SKV G S++A+LS Q PIPFNG + F F +D +V E +K
Sbjct: 3 LNKGTLEFDPTLVTDLISKVAGKSSIALRLSAQKPIPFNGEKVFFTTMDSEIDVVAESGKKT 62

Query: 65 GNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFLKHYADGFAKKMAEAFDIAAIHG 124
+ + + P+K Y AR++DEF YAS+E+++N L+ + DGFACK+A D+ A HG
Sbjct: 63 HGGVTLAPQTMVPIKVEYGARISDEFMYASDEEKINILQEFNDGFAKKVARGIDLMAFHG 122

Query: 125 LEPRTMT 131
+ PR T
Sbjct: 123 VNERLGT 129

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2300

A DNA sequence (GBSx2439) was identified in *S.agalactiae* <SEQ ID 7085> which encodes the amino acid sequence <SEQ ID 7086>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2301

A DNA sequence (GBSx2440) was identified in *S.agalactiae* <SEQ ID 7087> which encodes the amino acid sequence <SEQ ID 7088>. Analysis of this protein sequence reveals the following:

-2577-

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2227(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
10 vaccines or diagnostics.

Example 2302

A DNA sequence (GBSx2441) was identified in *S.agalactiae* <SEQ ID 7089> which encodes the amino acid sequence <SEQ ID 7090>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

15 Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.2948(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9319> which encodes amino acid sequence <SEQ ID 9320> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
 bacteriophage MML]
Identities = 84/238 (35%), Positives = 137/238 (57%), Gaps = 8/238 (3%)

30 Query: 1 MTLDKSSQAQKAGLILQEKIEDRLAIRNHSEMTYGLKKEYLKQWIPTVKDSTKRGYL 60
 +T++K + QA+ +A ++LQEKI +L+ + +T+ E+ + K W TVK+STK
Sbjct: 30 VTMEKKTTPQARNQAAILLQEKINKKLSTKQVESITFEEIYNLFYKSWAQTVKESTKHNC 89

35 Query: 61 VSDSHIATVLPDDTIINKLTKRDIRLIIDKLLKHNSYHVTHKCRKRLHAIFSYAIQMDYM 120
 D + V+P DTI+ L +R ++ I+K+++ N Y K R RL IF+YA+Q Y+
Sbjct: 90 SVDKKMKKEVIPSDTILANLDRRLQEAIEKIIESNGYITAKKVRHRLRGIFNYAVQYSYI 149

40 Query: 121 TSNPTENVLVP-KPK--DDYKPEKVLVLTSEV---YDLCNRMIDNDEQTLADIVLFMFL 174
 +N + +P KPK ++ + ++ +LT E+ D+ NR Q AD+VL + L
Sbjct: 150 ENNEVDYTTIPQPKPTLEELEKRNFLMQEIKALVDVLNRR--EYHQKYADMVLVLT 207

45 Query: 175 TGVRYGELSCLTYDKIDFENKEILINATYDFNTRXIITTKTKKSTRKISVSDNILDIV 232
 TG+RYGEL+ L IDFEN +I I +D + T KT S R I VS+++++ +
Sbjct: 208 TGMRYGELTALQLKNIDFENNKIEITGNFDSVNKIKTLPKTTNSIRTIKVSESIEAI 265

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2303

50 A DNA sequence (GBSx2444) was identified in *S.agalactiae* <SEQ ID 7091> which encodes the amino acid sequence <SEQ ID 7092>. Analysis of this protein sequence reveals the following:

Possible site: 50
>>> Seems to have no N-terminal signal sequence

-2578-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2518(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 4212:

10 Identities = 92/144 (63%), Positives = 118/144 (81%), Gaps = 1/144 (0%)

Query: 1 MPKYSLFELENGRRRLIASAGELQKGNELALPTQFMKFLYLASRYNESKKGKPEEIEKKQE 60
 +PKYSLFELENGR+R+LASAGELQKGNELALP++++ FLYLAS Y + KG PE+ E+KQ
 Sbjct: 1198 LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNEQKQL 1257

15 Query: 61 FVNQHVSYFDDILQLINDFSKRVLADANLEKINKLYQDNKENISVDELANNIINLFTFT 120
 FV QH Y D+I++ I++FSKRVLADANL+K+ Y +++ + E A NII+LFT T
 Sbjct: 1258 FVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRRDK-PIREQAENIIHLFTLT 1316

20 Query: 121 SLGAPAAFKFFDKIVDRKRYTSTQ 144
 +LGAPAAFK+FD +DRKRYTST+
 Sbjct: 1317 NLGAPAAFKYFDTTIDRKRYTSTK 1340

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 **Example 2304**

A DNA sequence (GBSx2445) was identified in *Sagalactiae* <SEQ ID 7093> which encodes the amino acid sequence <SEQ ID 7094>. This protein is predicted to be (). Analysis of this protein sequence reveals the following:

30 Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.57 Transmembrane 239 - 255 (236 - 256)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]
 Identities = 88/257 (34%), Positives = 129/257 (49%), Gaps = 11/257 (4%)

45 Query: 1 MARLGADFYSKLVTDLQKDGFKFYQQTGVFLKKDESQLESFLALADKRRLESPLIGD 60
 +A+ GA +Y L+ L+KDG Y++ G + D S+L+ + A KRR ++P IGD
 Sbjct: 61 LAKGGARYYKDLIHQLEKDGESDTGYKRVGAISIHTDASKLDKMEERAYKRREDAPEIGD 120

50 Query: 61 LQILNKSEANTHPPEL-DGYEQLLYASGGARVEGADLTRILLEAS---GVNVIKDEVHFP- 115
 + L+ SE FP L DGYE ++ SG ARV G L R LL A+ G VIK
 Sbjct: 121 ITRLSASETKKLFPIADGYES-VHISGAARVNGRALCRSLLSAABKRGATVIKGNASLL 179

55 Query: 116 ----TITDNGFRVQSIDFDKLVLASGAWLAKILDEHNYQVDVRPQKQQLRDYYFSNINTG 171
 T+T + D +++ +GAW +IL V QK Q+ + ++ +TG
 Sbjct: 180 FENGTVTGVQTDTKQFAADAVIVTAGAWANEILKPLGIHFQVSFQKAQIMHFEMTDADTG 239

 Query: 172 KYPVVMPEGELDIIPFDNGKVSVGASHENDMAF-DLNIDFKVLDKFEEQAIGYFFQLKKQ 230
 +PVVMP + I+ FDNG++ GA+HEND DL + + +A+ P L
 Sbjct: 240 SWPVVMPSPDQYILSFDNGRIVAGATHENDAGLDDLRTAGGQHEVLKALAVAPGLADA 299

 Query: 231 IRLLKRVEFVPIQVIFL 247
 + RV F P FL

-2579-

Sbjct: 300 AAVETRVGFRPFTPGFL 316

There is also homology to SEQ ID 2656.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2305

A DNA sequence (GBSx2446) was identified in *S.galactiae* <SEQ ID 7095> which encodes the amino acid sequence <SEQ ID 7096>. Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2572(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9315> which encodes amino acid sequence <SEQ ID 9316> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC00337 GB:AF008220 YtqI [Bacillus subtilis]
Identities = 119/256 (46%), Positives = 174/256 (67%), Gaps = 3/256 (1%)

Query: 6 QILDKIKEYDTIIHRHMRPDPDALGSQIGLRDIIRHNFPPKKVLATGFDEPTLAWIAKM 65
      +++ I YDTII+HRH+RPDEDA GSQ GL +I+R +P+K + A G EP+L+++ +
Sbjct: 4 ELIRTISLYDTIILHRHVRPDPDAYGSQCGLTEILRETYPEKNIFAVGTPEPSLSFLYSL 63

Query: 66 DQVTDQDYQGALVVVTDANTPRIDDERYKGDFLIKIDHHPNDEVYGDLSYVDTNASSA 125
      D+V ++ Y+GALV+V DTAN RIDD+RY G L+KIDHHPN++ YGDL +VD+ASS
Sbjct: 64 DEVDNETYEGALVIVCDTANQERIDDQRYPSGAKLMKIDHHPNEDFYGDLLWVDTSSASSV 123

Query: 126 SEIVTDFAL---SCDLLLSTSAARVLYNGIVGDTGRFLYPATTSKTLKIASKLREFDFDF 182
      SE++ + L L+T AA ++Y GIVGDTGRFL+P TT KTLK A +L ++ F
Sbjct: 124 SEMIYELYLEGKEHGWKLNTKAABLIYAGIVGDTGRFLFPNTTEKTLKYAGELIQYPFSS 183

Query: 183 SAMARQMDSFPFKIAKLQGFIFEQLKIDKNGAACVTLTQEDLKRFDVTDAAETAAIVGVPG 242
      S + Q+ + KL GFIF+ + + +NGAA V + ++ L++F T +E + +VG G
Sbjct: 184 SELFNQLYETKLNVVKLNGFIFQNVSLSENGAASVFIKDFTLEKFGTTASEASQLVGTILG 243

Query: 243 KIDIVESWAIFVKQSD 258
      I + +W FV++ D
Sbjct: 244 NISGIRAWVFFVEEDD 259

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7097> which encodes the amino acid sequence <SEQ ID 7098>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 180/256 (70%), Positives = 215/256 (83%)
Query: 4 FQILDKIKEYDTIIHRHMRPDPDALGSQIGLRDIIRHNFPPKKVLATGFDEPTLAWIA 63

```

-2580-

F+ ILDKIK + TIIHRH PDPDALGSQ GL++II NFP KKVL TGFDEP+LAWI+
 Sbjct: 5 FETILDKIKAHQTIIHRHQNPDALGSQAGLKEIIAQNFPDKKVLMTGFDEPSLAWIS 64

 Query: 64 KMDQVTDQDYQGALVVVTDANTPRIDDERYKKGDFLIKIDHHPNDEVYGDLSYVDTNAS 123
 5 +MDQVTD+DY+ ALV++TDTAN PRIDDERY G LIKIDHHPND+VYGD YVDT+AS
 Sbjct: 65 QMDQVTDKDYKEALVIIITDTANRPRIDDERYTLGKCLIKIDHHPNDVYGDYFYVDTAS 124

 Query: 124 SASEIVTDFALSCDLLLSTSAARVLYNGIVGDTGRFLYPATTSKTLKIASKLREFDFDFS 183
 SASEI+ DFA S +L LS AA++LY GIVGDTGRFLY +TTSKTL IAS+LR F+PDF+
 10 Sbjct: 125 SASEIADFAFSQNLTLSDKAAKLLYTGIVGDTGRFLYASTTSKTLASQLRHFEDFA 184

 Query: 184 AMARQMDSFPFKIAKLQGFIFEQLKIDKNGAACVTLTQEDLKRFDVTDAAATAIVGVPGK 243
 A++RQMDSFP KIAKLQ ++FE L ID++GAA V ++QE LK FDVT AE++AIV PGK
 15 Sbjct: 185 AISRQMDSFPLKIAKLQSYVFEHLTIDESGAAYVLVSQETLKHFDVTLAESSAIVCAPGK 244

 Query: 244 IDIVSWAIFVKQSDG 259
 ID V++WAIFV+ +DG
 Sbjct: 245 IDNVQAWAIFVELTDG 260

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2306

A DNA sequence (GBSx2447) was identified in *S. agalactiae* <SEQ ID 7099> which encodes the amino acid sequence <SEQ ID 7100>. Analysis of this protein sequence reveals the following:

25 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.1846(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:CAB42949 GB:AL049863 putative adenosine deaminase [Streptomyces
 coelicolor A3(2)]
 Identities = 123/343 (35%), Positives = 175/343 (50%), Gaps = 26/343 (7%)

 Query: 6 LKELAKAELHCHLDGSLSLPAIRKLANMADIILPSSDK-ELRKYVIAPAQTESLVDYLKT 64
 L+ L KA LH HLDG L + +LA LP++D EL + A + LV Y+ T
 40 Sbjct: 11 LRRLPKAVLHDHLDGGLRPATVVELARSVGHTLPTTDPDELAAWYYEAANGDLVRYIAT 70

 Query: 65 FEFIRPLLQTKALRFAAYDVARQAALENVIIYIEIRFAPELSMDKGLTASDTVLAVLEGL 124
 FE ++Q +E L AA + A + V+Y E+R+APEL+ GL+ + V V EGL
 45 Sbjct: 71 FEHTLAVMQNREGLLRAAEEYVLDLAADGVVYGEVRYAPELNTRGGLSMREVVETVQEGL 130

 Query: 125 ADAQKEFNIVAR-----ALVCGMRQSSHKTTKDIIKHIVDLA-----PKGLVGFDFAFAGDEF 175
 A + L+CGMR D ++ DLA G+VGFD AG E
 Sbjct: 131 ATGMAKAAAAGTPVRVGTLLCGMRMF-----DRVREAADLAVAFRDAGVVGFDIAGAED 184

 50 Query: 176 SYPTDSLVDLIQEVKRSYPMTHAGECGCAKHTADSLNL-GIKRMGHVTALT----- 227
 +P +D + ++R P T+HAGE I +L + G +R+GH +T
 Sbjct: 185 GFFPADHLDAFEHLRRENVPFTIHAGEAHLPSIHQALQVCGAQRIGHGVRTDDIPDLA 244

 Query: 228 -GORDLIKRFVEEDAVA-EMCLTSNLQTKAASSIQSFPYQELYDAGGKITINTDNRTVSD 285
 G+ + +V + +A EMC TSNLQT AA+SI P L D G ++T+NTDNR VS
 55 Sbjct: 245 AGKLGRLAANVRDRRIALEMCPTSNLQTGAATSAEHPITALKDLGPRVTLNTDNRLVSG 304

 Query: 286 TNLTKRYSLFVITYFGTKIEDFLVFNQNAVKASFTSDSEKDTLL 328
 T +T+E SL V G +ED NA+K++F E+ L+
 60 Sbjct: 305 TTMTREMSLLVEQAGWSVEDLRITVTNALKSFAFVPFDERTALI 347

-2581-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2307

- 5 A DNA sequence (GBSx2448) was identified in *S.agalactiae* <SEQ ID 7101> which encodes the amino acid sequence <SEQ ID 7102>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 A related GBS nucleic acid sequence <SEQ ID 9639> which encodes amino acid sequence <SEQ ID 9640> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:CAB13290 GB:Z99111 similar to sulfite reductase [Bacillus subtilis]
Identities = 63/146 (43%), Positives = 87/146 (59%), Gaps = 1/146 (0%)
20
Query: 5 MALAKIVYASMTGNTTEIADIVADKLRDLGLDVEVEECTMVDAAD-FEDADIAIVATYTYT 63
MA +VYA+M+GNTE +AD++ L++ +V+ E +D A F D D I+ TYT+
Sbjct: 1 MAKILLVYATMSGNTEAMADLIEKGLQEALAEVDRFEAMDIDDAQLFTDYYDHVIMGTYTW 60
25
Query: 64 GDGDLPEIVDFYEDLAEVDLSGKVYGVVSGDFTFYDYFCKSVDEFEAQFALTGAQKGAD 123
GDGDLPE +D ED+ E+D SGK V GSGDT Y++FC +VD EA+ G
Sbjct: 61 GDGDLPEFLDLVEDMEEIDFSGKTCVFGSGDTAYEFFCGAVDTLEAKIKERGGDIVLP 120
30
Query: 124 CVKVDLAAEDEDIENLEAFAEETASK 149
VK++ E E+ E L F + A K
Sbjct: 121 SVKIENNPGESEEEELINFRQFAKK 146

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7103> which encodes the amino acid sequence <SEQ ID 7104>. Analysis of this protein sequence reveals the following:

- 35 Possible site: 14
>>> Seems to have no N-terminal signal sequence
40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1641(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 116/147 (78%), Positives = 136/147 (91%)
45
Query: 5 MALAKIVYASMTGNTTEIADIVADKLRDLGLDVEVEECTMVDAADFEDADIAIVATYTYG 64
MALAKIVYASMTGNTTEIADIVA+KL++LG DV+++ECT VDA++FE+ADIA+VATYTYG
Sbjct: 1 MALAKIVYASMTGNTTEIADIVANKLQELGHDVDIDECTTVDASEFENADIAVATYTYG 60
50
Query: 65 DGDLPDEIVDFYEDLAEVDLSGKVYGVVSGDFTFYDYFCKSVDEFEAQFALTGAQKGAD 124
DGDLPDEIVDFYEDL ++DL GK+YGVVSGDFTFYDYFCKSVDF QFALTGA KGA+
Sbjct: 61 DGDLPDEIVDFYEDLQDLGLEKVIYGVVSGDFTFYDYFCKSVDDFSEQFALTGAIKGAEP 120
55
Query: 125 VKVDLAAEDEDIENLEAFAEETASKLN 151
VKVDLAAEEDI+ LEAFAE+++ +N
Sbjct: 121 VKVDLAAEDEDIDLEAFAEQLSQAVN 147

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2308

- 5 A DNA sequence (GBSx2449) was identified in *S.agalactiae* <SEQ ID 7105> which encodes the amino acid sequence <SEQ ID 7106>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3568(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB98234 GB:U67480 chorismate mutase/prephenate dehydratase
(pheA) [Methanococcus jannaschii]
Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

20 Query: 2 ELEEEIRQBEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDDNNRENEVLNNVLKKVQN 61
+L EIR++IDEID +++ L+ R L +V K + +P+ D RE + + + K +
Sbjct: 4 KLAIEIRKKIDEIDNKILKLAERNLAKDVAEIKNQLGIPINDFEREKYIYDRIRKLCKE 63

25 Query: 62 HQFDDVIRATFKDIMTE-SRVYQKE 85
H D+ I I+ E ++ QK+
Sbjct: 64 HNVDENIGIKIFQILIEHNKALQKQ 88

There is also homology to SEQ ID 1568.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2309

A DNA sequence (GBSx2450) was identified in *S.agalactiae* <SEQ ID 7107> which encodes the amino acid sequence <SEQ ID 7108>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

- 35 Possible site: 23
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1828(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 45 >GP:AAC34413 GB:AF158600 putative minor structural protein
[Streptococcus thermophilus bacteriophage Sfil1]
Identities = 39/65 (60%), Positives = 54/65 (83%)

50 Query: 1 MEVETDSQEVLMSGLKDLKAHAYPAITYEVDGVDLELGDVVRIQDDGYEPPLILTARV 60
ME++TDS++VL+ST L++L+ YPAITYEVDG++DL++GD V+IQD G+ P L+L ARV
Sbjct: 707 MEIDTDSQVLISTALRNLRKFCYPAITYEVDGFLDLDIGDTVKIQDTGFSPMLMLEARV 766

Query: 61 VEQDI 65
EQ I
Sbjct: 767 SEQQI 771

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2310

A DNA sequence (GBSx2451) was identified in *S.agalactiae* <SEQ ID 7109> which encodes the amino acid sequence <SEQ ID 7110>. This protein is predicted to be phosphomethylpyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

```

10 Possible site: 45
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.2051(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:AAC22074 GB:U32725 phosphomethylpyrimidine kinase (thiD)
    [Haemophilus influenzae Rd]
    Identities = 29/78 (37%), Positives = 48/78 (61%), Gaps = 2/78 (2%)

    Query: 4  RNVLAISGNDIFSGGGLHADLATYVVNKLHGFVAVTCLTAMSDKG-FEVIPIEASILKQQ 62
              + VL I+G+D  G G+ ADL T+ + + G  A+T +TA + G F++ PI  ++ Q
    Sbjct: 5  KQVLTIAGSDSGGAGIQADLKTFFQMRGVFGTSAITAVTAQNTLGVFDIHPIPLKTIQQA 64

25 Query: 63  LESLK-DVEFGSIKGLLL 79
              LE++K D + S K+G+L
    Sbjct: 65  LEAVKNDFQIASCKIGML 82

```

30 There is also homology to SEQ ID 4408.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2311

35 A DNA sequence (GBSx2452) was identified in *S.agalactiae* <SEQ ID 7111> which encodes the amino acid sequence <SEQ ID 7112>. Analysis of this protein sequence reveals the following:

```

    Possible site: 57
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -7.43    Transmembrane  109 - 125 ( 102 - 129)
        INTEGRAL    Likelihood = -1.28    Transmembrane   84 - 100 ( 84 - 100)
40
    ----- Final Results -----
                bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50 >GP:CAA22372 GB:AL034446 putative transmembrane protein
    [Streptomyces coelicolor A3(2)]
    Identities = 25/93 (26%), Positives = 43/93 (45%), Gaps = 1/93 (1%)

    Query: 62  SASVEILCRGWLLPVSATKYSKIVSVSISIFFGLLHSANNHVSLSIFNLCL-FGLFLS 120
              +A+ E++ RG L  +      +++ ++ + FGL+H N  +L      + + G L+
    Sbjct: 143  AATEVVVFRGVLFRIIEEHIGTYLALGLTGLVFLMHLNEDATLWGALATAIEAGFMLA 202

```

-2584-

Query: 121 LYVILKGNIWGACGIHGAWNCVQGSVFGIEVSG 153
 N+W G+H WN G VF VSG
 Sbjct: 203 AAYAATRNLWLTIGVHFGWNFAAGGVFSTVVS 235

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2312

A DNA sequence (GBSx2453) was identified in *S.agalactiae* <SEQ ID 7113> which encodes the amino acid sequence <SEQ ID 7114>. This protein is predicted to be pppL protein. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5796(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA10712 GB:AJ132604 pppL protein [Lactococcus lactis]
 Identities = 38/64 (59%), Positives = 51/64 (79%)

Query: 1 MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
 ME S+L+DIG +RS NQD++ + N+AG L +LADGMGGH+AGN+AS++TV DLG W+
 Sbjct: 1 MEYSILSDIGSKRSTNQDYVGTYVNRAGYQLFLLADGMGGHKAGNVASKLTVEDLGKLWS 60

Query: 61 ETDF 64
 ET F
 Sbjct: 61 ETFF 64

There is also homology to SEQ ID 3022:

Identities = 58/74 (78%), Positives = 69/74 (92%)

Query: 1 MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
 M+ISL TDIGQ+RSNNQDFIN+F+NK G+ L+ILADGMGGHRAGNIASEMTVTDLG +W
 Sbjct: 1 MKISLKTIDIGKRSNNQDFINKFDNKKGITLVILADGMGGHRAGNIASEMTVTDLGREWV 60

Query: 61 ETDFSELSEIRDWM 74
 +TDF+ELS+IRDW+
 Sbjct: 61 KTDFTELSQIRDWL 74

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2313

A DNA sequence (GBSx2454) was identified in *S.agalactiae* <SEQ ID 7115> which encodes the amino acid sequence <SEQ ID 7116>. This protein is predicted to be sunL protein. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

-2585-

bacterial cytoplasm --- Certainty=0.1631(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA10711 GB:AJ132604 sunL protein [Lactococcus lactis]
 Identities = 48/81 (59%), Positives = 67/81 (82%)

10 Query: 1 MSILSSVCQTLRKGGIITYSTCTIFEEENFQVIRKFLNHPNFEQVELSHTQEDIVKRG 60
 + IL+S ++L+K GI+ YSTCTIF+EENF V+ +FLENHPNFEQVE+S+ + +++K GC
 Sbjct: 342 LEILNSASKSLKKSGIMVYSTCTIFDEENFDVVHEFLENHPNFEQVEISNEKPEVIKEGC 401

Query: 61 ISISPEQYHTDGGFFIGQVKRI 81
 + I+PE YHTDGGFI + K+I
 15 Sbjct: 402 LFITPEMYHTDGGFFIAKFKKI 422

There is also homology to SEQ ID 3018:

Identities = 64/82 (78%), Positives = 74/82 (90%)

20 Query: 1 MSILSSVCQTLRKGGIITYSTCTIFEEENFQVIRKFLNHPNFEQVELSHTQEDIVKRG 60
 + ILSSVCQTLRKGGIITYSTCTIF+EEN QVIE FL++HPNFEQV+L+HTQ DIVK G
 Sbjct: 359 LEILSSVCQTLRKGGIITYSTCTIFDEENRQVIEAFLQSHPNFEQVKNHTQADIVKDG 418

Query: 61 ISISPEQYHTDGGFFIGQVKRI 82
 + I+PEQY TDGFFIGQV+R+L
 25 Sbjct: 419 LIITPEQYQTDGFFIGQVRRVL 440

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 **Example 2314**

A DNA sequence (GBSx2455) was identified in *Sagalactiae* <SEQ ID 7117> which encodes the amino acid sequence <SEQ ID 7118>. This protein is predicted to be PTS permease for mannose subunit IIPMan. Analysis of this protein sequence reveals the following:

Possible site: 53

35 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.18 Transmembrane 32 - 48 (30 - 58)
 INTEGRAL Likelihood = -8.07 Transmembrane 127 - 143 (122 - 146)
 INTEGRAL Likelihood = -2.07 Transmembrane 56 - 72 (56 - 72)
 40 INTEGRAL Likelihood = -1.44 Transmembrane 87 - 103 (86 - 103)
 INTEGRAL Likelihood = -0.53 Transmembrane 105 - 121 (105 - 121)

----- Final Results -----

bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF81084 GB:AF228498 AgaW [Escherichia coli]
 Identities = 38/122 (31%), Positives = 68/122 (55%), Gaps = 7/122 (5%)

50 Query: 25 KVPETKSIIRLTALAFVCSILVVELVSMRELISSISFIGILVSGPVSFVHHIPQNL 84
 ++P T + L A +L L+++ +F+ I G+ + + +PQ L+
 Sbjct: 126 RMPRTFILAALNACNYLA-----LLALGNFYFLCAFLPIYFGAEHAKTIIDVLPQR 178

55 Query: 85 NGLSAAGGLLPVAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAAALGAVICVISS 144
 +GL AGG++PA+GFA+L+K++ N +++LGFV A+LKL P+A+A + +I
 Sbjct: 179 DGLGVAGGIMPATGFAVLLKIMMKNVYIPYFILGFVAAWLKLPLVLAIAACPALAMALIDL 238

Query: 145 QR 146

-2586-

R
Sbjct: 239 LR 240

There is also homology to SEQ ID 1636:

5 Identities = 104/109 (95%), Positives = 108/109 (98%)

Query: 56 LISSISFIGILVGSGPVNSFVHHIPQNLNGLSAAGGLLPAVGFAMLMKLLWTKLAVFY 115
 +I+SISFIGILVGSGPVN+VH HIPQNLNGLSAAGGLLPAVGFAMLMKLLWTKLAVFY

10 Sbjct: 149 IIASISFIGILVGSGPVNAFVEHIPQNLNGLSAAGGLLPAVGFAMLMKLLWTKLAVFY 208

Query: 116 LLGFVLTAYLKLPVAVAAALGAVICVISSQDIELDAITRGAISKQTTF 164
 LLGFVLTAYLKLPVAVAAALGAVICVISSQD+ELDAITRGAISKQTTF

 Sbjct: 209 LLGFVLTAYLKLPVAVAAALGAVICVISSQDLELDAITRGAISKQTTF 257

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2315

A DNA sequence (GBSx2456) was identified in *Sagalactiae* <SEQ ID 7119> which encodes the amino acid sequence <SEQ ID 7120>. Analysis of this protein sequence reveals the following:

20 Possible site: 50
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -8.12	Transmembrane	121 - 137 (118 - 144)
INTEGRAL	Likelihood = -5.52	Transmembrane	91 - 107 (89 - 111)
INTEGRAL	Likelihood = -5.20	Transmembrane	166 - 182 (162 - 192)

25 ----- Final Results -----

bacterial membrane	---	Certainty=0.4248(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15963 GB:Z99124 phosphotransferase system (PTS)
 beta-glucoside-specific enzyme IIABC component [Bacillus subtilis]

35 Identities = 76/201 (37%), Positives = 122/201 (59%), Gaps = 3/201 (1%)

Query: 1 MIKALLALLLVFKILTPSSQTYILLNLFADGVFVFLPILIIAIAAQKLNKLPILALGTVV 60
 MIK L+AL + F + SQ +++L DG FYFLP+L+A++AA+K +NP +A

 Sbjct: 121 MIKGLVALAVTFGWMAEKSQVHVILTAVGDGAFYFLPLLAMSAARKFGSNPYVAAAI 180

40 Query: 61 MLLHPNWNANLVASGKPVSLFHTIPFTLTNYASSVIPIIILICVQAYIEKYLKQIIPKSLR 120
 +LHP+ L+ +GKP+S F +P T Y+S+VIPI+L I + +Y+EK++ + SL+

 Sbjct: 181 AILHPDLTALLGAGKPIS-FIGLPVTAATYSSTVIPILLSIWIASVVEKNIDRFTHASLK 239

45 Query: 121 LVLVPMLIFLSMGILSFSILGPMGTIAGQYLAVIFTFLSKYASW-APAFVLVGAFAFILIM 179
 L++VP L + L+ +GP+G I G+YL+ +L +A A FL G F+ ++IM

 Sbjct: 240 LIVVPTFTLLIVVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMIFLAGTFS-LIIM 298

 Query: 180 FGVHSGIAALGITQLAKLGVD 200
 G+H + I +A+ G D

50 Sbjct: 299 TGMHYAFVPIMINNIAQNGHD 319

There is also homology to SEQ ID 2884.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2587-

Example 2316

A DNA sequence (GBSx2457) was identified in *S.agalactiae* <SEQ ID 7121> which encodes the amino acid sequence <SEQ ID 7122>. This protein is predicted to be glucose kinase. Analysis of this protein sequence reveals the following:

5 Possible site: 54
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1180(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB14416 GB:Z99116 glucose kinase [Bacillus subtilis]
 Identities = 32/57 (56%), Positives = 41/57 (71%)

 Query: 1 MVIGGGVSAAGEFLRSRVEKYFVTFAPFPQVKKSTKIKIAELGNDAGIIGAASLANQ 57
 +V+GGGVS AGE LRS+VEK F AFP+ ++ I IA LGNDAG+IG A +A +
 20 Sbjct: 258 IVLGGGVSFRAGELLRSKVEKIFRKCAFPRAQAADISIAALGNDAGVIGGAWTAKNE 314

There is also homology to SEQ ID 198. An alignment of the GAS and GBS proteins is shown below:

 Identities = 50/56 (89%), Positives = 53/56 (94%)

25 Query: 1 MVIGGGVSAAGEFLRSRVEKYFVTFAPFPQVKKSTKIKIAELGNDAGIIGAASLANQ 56
 +VIGGGVSAAGEFLRSR+EKYFVTF FPOV+ STKIKIAELGNDAGIIGAASLA Q
 Sbjct: 264 VVIGGGVSAAGEFLRSRIEKYFVTFTFPQVRYSTKIKIAELGNDAGIIGAASLARQ 319

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2317

A DNA sequence (GBSx2458) was identified in *S.agalactiae* <SEQ ID 7123> which encodes the amino acid sequence <SEQ ID 7124>. Analysis of this protein sequence reveals the following:

35 Possible site: 19
 >>> Seems to have a cleavable N-term signal seq.

 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAB14385 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 37/86 (43%), Positives = 51/86 (59%)

 Query: 3 MSVILIIVILLAFVAWASWNYWRVRAAKFLDNESFOKEMSRGQLIDIREAGAFHRKHIL 62
 MS +++++I AF+ + +Y +R K L E F+ + QLID+RE F HIL
 Sbjct: 1 MSNMIVLIIFPAFIYMIASVYVQQRIMKTLTEEBFRAGYRKAQLIDVREPNEFEGGHIL 60

50 Query: 63 GARNIPASQFKVALSALRKDKPVLLY 88
 GARNIP SQ K + +R DKPV LY
 Sbjct: 61 GARNIPLSQLKQKNEIRTDKPVYLY 86

There is also homology to SEQ ID 202. An alignment of the GAS and GBS proteins is shown below:

55 Identities = 51/108 (47%), Positives = 70/108 (64%)

-2588-

Query: 1 MDMSVILLIIVILLAFVAVASWNYWRVRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKH 60
 M +++ ++L+ V + +WNY+ R+ AK +DNE+F+ M +GQLID+RE AF KH
 Sbjct: 1 MSPITLILWLLLVGIVGYTWNYSFRKMAKQVDNETFKDVMRQQLIDLREPAAFRTKH 60

5 Query: 61 ILGARNIPASQFKVALSALRKDKPVLLYDASRGQSIPRIVLLLRKERF 108
 ILGARN PA QF A+ LRKDKPVL+Y+ R Q V L+K F
 Sbjct: 61 ILGARNFPAQQFDAAIKGLRKDKPVLIIYENMRPQYRVPVAVKKLKKAGF 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2318

A DNA sequence (GBSx2459) was identified in *S.agalactiae* <SEQ ID 7125> which encodes the amino acid sequence <SEQ ID 7126>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

15 Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.1892 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2319

A DNA sequence (GBSx2460) was identified in *S.agalactiae* <SEQ ID 7127> which encodes the amino acid sequence <SEQ ID 7128>. Analysis of this protein sequence reveals the following:

30 Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.3522 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2320

A DNA sequence (GBSx2461) was identified in *S.agalactiae* <SEQ ID 7129> which encodes the amino acid sequence <SEQ ID 7130>. Analysis of this protein sequence reveals the following:

45 Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2770 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-2589-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAB18708 GB:U38906 ORF33 [Bacteriophage rlt]
Identities = 56/85 (65%), Positives = 66/85 (76%), Gaps = 1/85 (1%)

Query: 1 MTNFATTDDVILLWRQLSVDEIKRAEALLETVSDTLRLLEASKVGKNLDEMILETP-YFAT 59
M FAT DD+ +LWR L DE +RAE LLE VSD+LR EA KVG++L MI E P YFA+
10 Sbjct: 1 MNPFATVDDLITMLWRPLKGDEKERAELILEIVSDSLREEADKVGRLDYAMIAEKPSYFAS 60

Query: 60 VLKSVTVDIVARTLMTATQGEPMQ 84
V+KSVTVDIVARTLMT+T EPM+Q
Sbjct: 61 VVKSVTVDIVARTLMTSTDQEPMTQ 85

15 There is also homology to SEQ ID 1432.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2321

20 A DNA sequence (GBSx2462) was identified in *S.agalactiae* <SEQ ID 7131> which encodes the amino acid sequence <SEQ ID 7132>. This protein is predicted to be regulatory protein TypA (typA). Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2238(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06351 GB:AP001516 GTP-binding protein TypA/BipA (tyrosine phosphorylated protein A) [Bacillus halodurans]
Identities = 175/237 (73%), Positives = 204/237 (85%), Gaps = 1/237 (0%)

35 Query: 1 MEDIFVGETVTPTDAIEPLPVLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 60
ME+I VGETV P D +PLP+LRIDEPTLQMTFLVNNSPFAGREGK +TSRK+EERL AE
Sbjct: 281 MEEINVGETVCPVDHQPILRLIDEPTLQMTFLVNNSPFAGREGKHVTSRKLEERLRAE 340

40 Query: 61 LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIEIETMRREGYELQVSRPEVIIKEIDGVQCE 120
L+TDVSLRV+ TDSPD W VSGRGELHLSILIE MRREGYELQVS+PEVII+EIDGVQCE
Sbjct: 341 LETDVSLRVNTDSPDMWVSGRGELHLSILIEIETMRREGYELQVSKPEVIIKEIDGVQCE 400

45 Query: 121 PFERVQIDTPEEYQGAIIQSLSERKGMMLDMQVGNQQLRIFLIPARGLIGYSTEFLSM 180
P ERVQID PEEY GA+++SL ERKG+ML+M G+GQ RL F++PARGLIGY+TEFLS
Sbjct: 401 PVERVQIDVPEEYTGAVMESLGERKGEMLNMNTNGSGQVRLEFMPARGLIGYTFEFLSQ 460

Query: 181 TRGYGIMNHTFDQYLPVVQGEIGGRHGRGALVSIENGKATTYSIMRIEERGNLSFVNP 237
TRGYGI+NH+FD Y PV G++GGR +G LVS+E GKAT Y I+++E+RG + FV P
Sbjct: 461 TRGYGIINHSFDSYQVTPGQVGGRRQGVLSMETGKATQVGIIQVEDRGTI-FVEP 516

50

There is also homology to SEQ ID 206. An alignment of the GAS and GBS proteins is shown below:

Identities = 228/237 (96%), Positives = 233/237 (98%), Gaps = 1/237 (0%)

55 Query: 1 MEDIFVGETVTPTDAIEPLPVLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 60
MEDIFVGET+TPTD +E LP+LRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE
Sbjct: 284 MEDIFVGETITPTDCVEALPILRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 343

Query: 61 LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIEIETMRREGYELQVSRPEVIIKEIDGVQCE 120

-2590-

LQTDVSLRVDPTDSPDKWTVSGRGELHLSILLETMRREGYELQVSRPEVIIKEIDGV+CE
 Sbjct: 344 LQTDVSLRVDPTDSPDKWTVSGRGELHLSILLETMRREGYELQVSRPEVIIKEIDGVKCE 403
 Query: 121 PFERVQIDTPEEYQGAIQSLSERKGDMLDMQVNGQTRLIFLIPARGLIGYSTEFLSM 180
 5 PFERVQIDTPEEYQGAIQSLSERKGDMLDMQVNGQTRLIFLIPARGLIGYSTEFLSM
 Sbjct: 404 PFERVQIDTPEEYQGAIQSLSERKGDMLDMQVNGQTRLIFLIPARGLIGYSTEFLSM 463
 Query: 181 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIBERGNSLFSVNP 237
 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIBERG + FVNP
 10 Sbjct: 464 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIBERGII-FVNP 519

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2322

15 A DNA sequence (GBSx2464) was identified in *S.agalactiae* <SEQ ID 7133> which encodes the amino acid sequence <SEQ ID 7134>. This protein is predicted to be pseudouridine synthase family 1 protein (rhuB). Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1950(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14248 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 59/105 (56%), Positives = 85/105 (80%)
 30 Query: 5 VKERIYPVGRLDWDTTGLLILTNDGDFTDKMIHPRNEIDKVYLARVKG IATKENLRPLTR 64
 + +RIYP+GRLD+DT+GLL+LTNDG+F +K++HP+ EIDK Y+A+VKG I KE LR L R
 Sbjct: 91 IPQRIYPIGRLDYDTSGLLLTNDGDFANKLMHPKYEIDKTYVAKVKGIPPKELLRLKLER 150
 Query: 65 GVVIDGKKTKPARYTIIKVDHEKNRSVVELTIEGRNHQVKKMFE 109
 35 G+ ++ KT PA+ ++ +D +K S+++LTIHEGRN QV++MFE
 Sbjct: 151 GIRLEEGKTAPAKAKLLSLDKKKQTSIIQLTIEGRNRQVRRMFE 195

There is also homology to SEQ ID 4728:

Identities = 96/109 (88%), Positives = 106/109 (97%)
 40 Query: 1 MLPQVKERIYPVGRLDWDTTGLLILTNDGDFTDKMIHPRNEIDKVYLARVKG IATKENLR 60
 +LPQVKERIYPVGRLDWDT+G+LILTNDGDFTD MIHPRNEIDKVYLARVKG IATKENLR
 Sbjct: 94 LLPQVKERIYPVGRLDWDTSGVLILTNDGDFDTMIHPRNEIDKVYLARVKG IATKENLR 153
 Query: 61 PLTRGVVIDGKKTKPARYTIIKVDHEKNRSVVELTIEGRNHQVKKMFE 109
 45 PLTRG+VIDGKKTKPARY I++V+ +K+RS+VELTIEGRNHQVKKMFE
 Sbjct: 154 PLTRGIVIDGKKTKPARYNIVRVEADKRSIVELTIEGRNHQVKKMFE 202

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2323

A DNA sequence (GBSx2466) was identified in *S.agalactiae* <SEQ ID 7135> which encodes the amino acid sequence <SEQ ID 7136>. This protein is predicted to be L-ribulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

-2591-

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2827(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase
 [Bacillus stearothermophilus]
 Identities = 68/103 (66%), Positives = 82/103 (79%)

15 Query: 2 QEMRERVCEANKSLP VHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDL 61
 +E+++ V EAN LP + LV FTWGNVS +DRE GL+VIKPSGV YD+LT ++MVV DL
 Sbjct: 3 EELKQAVLEANLQLPQYRLVTFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDLT 62

 Query: 62 GNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHSTHSTEAVGWAQ 104
 GN+VEGDL PSSD PTH+ LYK +P +GGIVHSTHST A WAQ

20 Sbjct: 63 GNVVEGDLKPSSDTPTHLWLYKQFPGIGGIVHSTHSTWATVWAQ 105

There is also homology to SEQ ID 4600:

 Identities = 93/103 (90%), Positives = 96/103 (92%)

25 Query: 2 QEMRERVCEANKSLP VHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDL 61
 QEMRERVCAANKSLP H LVKFTWGNVSEV RE G IVIKPSGVDYD LTPENMVVTDL+
 Sbjct: 6 QEMRERVCAANKSLPQHGLVKFTWGNVSEVCRELGRIVIKPSGVDYDLLTPENMVVTDL 65

 Query: 62 GNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHSTHSTEAVGWAQ 104
 GN+VEGDLNPSSDLPTHV+LYKAWPEVGGIVHSTHSTEAVGWAQ

30 Sbjct: 66 GNVVEGDLNPSSDLPTHVELYKAWPEVGGIVHSTHSTEAVGWAQ 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 2324

A DNA sequence (GBSx2467) was identified in *S.agalactiae* <SEQ ID 7137> which encodes the amino acid sequence <SEQ ID 7138>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3452(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45

The protein has homology with the following sequences in the GENPEPT database.

 >GP:AAG05712 GB:AE004658 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 141/200 (70%), Positives = 162/200 (80%), Gaps = 1/200 (0%)

50 Query: 10 LSLGTDYETLANRFRPIFREISAGNVEREKARALPYEPIEWLKKAGFGAVRVPSEYGGAG 69
 LS G DYE LA RFRPIF I+ G VERE+ R LP+E I WLK+AGFGAVRVP E+GGAG
 Sbjct: 14 LSEGADYELLAQRFRPIFARIAEGAVERERQRELPHAEIWLKQAGFGAVRVPREHGGAG 73

 Query: 70 ASIGQLFQLLIELAEADSNIQALRAHFVVEDRLNAPPGVDRDTWFARFVAGDLVGNW 129
 AS+ QL QLLIELAEADSNI QALR HFAFVEDRLNA PG RD W RFV GDLVG W

55 Sbjct: 74 ASLPQLVQLLIELAEADSNIQALRGHFAFVEDRLNAEPGPGRDRLRRFVEGDLVGCAW 133

 Query: 130 TEVGTVKIGDVITKVSAGGDG-FVLNGTKFYSTGSIFADWIDVYAQRADNGADVIAVNA 188
 TEVG+V++G+V+T+VS + DG +V+NG+K+YSTGS+F+DWID+YAQR D GADVIA +

-2592-

Sbjct: 134 TEVGSVRLGEVLTRVSRKDDGRWVNGSKYYSTGSLFSDWIDLQAQRDDTGADVIAAIRT 193

Query: 189 RHAGVRHSDDWDGFGQRTTG 208
GVR SDDWDGFGQRTTG

5 Sbjct: 194 DQPGVRQSDDWDGFGQRTTG 213

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 2325

A DNA sequence (GBSx2468) was identified in *S.agalactiae* <SEQ ID 7139> which encodes the amino acid sequence <SEQ ID 7140>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1919(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

20

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2326

A DNA sequence (GBSx2474) was identified in *S.agalactiae* <SEQ ID 7141> which encodes the amino acid sequence <SEQ ID 7142>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

30

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2978(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

35

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 2327

A DNA sequence (GBSx2476) was identified in *S.agalactiae* <SEQ ID 7143> which encodes the amino acid sequence <SEQ ID 7144>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5402(Affirmative) < succ>

-2593-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2328

- 10 A DNA sequence (GBSx2477) was identified in *S.agalactiae* <SEQ ID 7145> which encodes the amino acid sequence <SEQ ID 7146>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2755 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 190/247 (76%), Positives = 225/247 (90%)

- 25 Query: 1 MELGQLFHHLGSEITLMQRSERLLKEYDPEISESVEKALIEQGGINLVKGATFERVEQSGE 60
 MELGQLFH+LGSE+TL+QRSERLLKEYDPEISESVEK+L+EQGINLVKGAT+ER+EQ+G+
 Sbjct: 262 MELGQLFHNLGSEVTILQRSERLLKEYDPEISESVEKSLVEQGGINLVKGATYERIEQNGD 321
- 30 Query: 61 IKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLAAAGVETGKNNELINDFGQTSNEK 120
 IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAGVE G EI+I+D+ +T+N +
 Sbjct: 322 IKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAGVEIGSRGEIIDDYSRTTNTR 381
- 35 Query: 121 IYAAGDVTLGPFQFVYVAAYEGGIITDNAIGGLNKKIDLSVVPVAVTFTNPTVATVGLTEEQ 180
 IYAAGDVTLGPFQFVYVAAY+GG+ NAIGGLNKK++L VVP VTFT P +ATVGLTE+Q
 Sbjct: 382 IYAAGDVTLGPFQFVYVAAYQGGAAPNAIGGLNKKLNLEVVPGVTFTAPAIATVGLTEEQ 441
- 40 Query: 181 AKEGYDVKTSVLPLGAVPRALVNRETTGVFKLVADAETLKVILGVHIVSENAGDVIYAAS 240
 AKE GY+VKTSVLPL AVPRA+VNRETTGVFKLVAD++T+KVLG H+V+ENAGDVIYAA+
 Sbjct: 442 AKENGVEVKTSLPLDAVPRALVNRETTGVFKLVADSKIMKVLGAHVVAENAGDVIYAAT 501
- 40 Query: 241 LAVKFGL 247
 LAVKFGL
 Sbjct: 502 LAVKFGL 508

There is also homology to SEQ ID 1820.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2329

A DNA sequence (GBSx2478) was identified in *S.agalactiae* <SEQ ID 7147> which encodes the amino acid sequence <SEQ ID 7148>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2594-

```

bacterial cytoplasm --- Certainty=0.3642(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2330

10 A DNA sequence (GBSx2479) was identified in *S.agalactiae* <SEQ ID 7149> which encodes the amino acid sequence <SEQ ID 7150>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

```

15

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1936(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

20

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2331

25 A DNA sequence (GBSx2480) was identified in *S.agalactiae* <SEQ ID 7151> which encodes the amino acid sequence <SEQ ID 7152>. This protein is predicted to be Nra. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence

```

30

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1510(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35

A related GBS nucleic acid sequence <SEQ ID 9383> which encodes amino acid sequence <SEQ ID 9384> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7153> which encodes the amino acid sequence <SEQ ID 7154>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.64 Transmembrane 22 - 38 ( 22 - 38)

```

45

```

----- Final Results -----
bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2595-

An alignment of the GAS and GBS proteins is shown below.

Identities = 42/157 (26%), Positives = 78/157 (48%), Gaps = 2/157 (1%)

```

5  Query: 71  LLGREFIDSQHFKDINAYFLRHFICYCYFIPDFYFLNLSRLSY--SKDLYHLLDKGLAD 128
      LLG  ++S  FK I   F R FI      +PD +  +  R      +K  Y+ L   + +
      Sbjct: 8  LLGNILNLSLPFKRILVSFSRLFISNLQVLLPDIHLFHYLRRQQRNKSFYNTLKTIVEE 67

10 Query: 129 IFNLKGGNLTFSKHETVLLTMQLSNLIETFLAPLSVYVVISSNIRLQTYQVMLNQYFTSK 188
      + +G      +  +L T+QL  L++T+L P+ VY+++++  L      L+ YF
      Sbjct: 68 WMSAEGIVGKLPSYHLLLFITIQLEELLKTYLPPIPVYLLTNNTAALDLMTNALSIFYFPPA 127

      Query: 189 IAEFFVFNYYQTQIDEKLLKKADIIIAERRYISSLKN 225
      IA  VN +      + + +K  +IIA+R+Y++ +++
15  Sbjct: 128 IATVMPVNVEIIPFKDIVKEKQSVIIADRQYLNLIQH 164

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2332

20 A DNA sequence (GBSx2481) was identified in *S.agalactiae* <SEQ ID 7155> which encodes the amino acid sequence <SEQ ID 7156>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

```

25  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1383 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2333

35 A DNA sequence (GBSx2482) was identified in *S.agalactiae* <SEQ ID 7157> which encodes the amino acid sequence <SEQ ID 7158>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

```

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4145 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2596-

Example 2334

A DNA sequence (GBSx2484) was identified in *S.agalactiae* <SEQ ID 7159> which encodes the amino acid sequence <SEQ ID 7160>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 57
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.02    Transmembrane    34 - 50 ( 34 - 50)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2335

A DNA sequence (GBSx2485) was identified in *S.agalactiae* <SEQ ID 7161> which encodes the amino acid sequence <SEQ ID 7162>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 49
   >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30   >GP:CAB52002 GB:AL109663 hypothetical protein [Streptomyces
      coelicolor A3(2)]
      Identities = 61/141 (43%), Positives = 86/141 (60%), Gaps = 2/141 (1%)

      Query: 3   TYFDNPLKTNQAYADLHCTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRNAG 62
                T D ++ N+ YA          + +P +VA+V CMD+RL + ALGL LGD H +RNAG
35   Sbjct: 5   TVTDRLEVEANERYAAAFADPGMDARPVQRVAVVACMDARLDLHAALGLKLGDCHTIRNAG 64

      Query: 63   GRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFTNEAFAAQLQRDLGVDMHGHDFLP 122
                G VTDDV+RSL ISQ+ LGTR + ++HHT CG +T T E F L+ ++G
40   Sbjct: 65   GVVTTDDVIRSLTISQRALGTRSVALLHHTGCGMETITEE-FRHDLELEVQ-QRPAAWAVEA 122

      Query: 123  FNDIEESVREDVAKLHASPFLL 143
                F D ++ VR+ + ++ SPFL
      Sbjct: 123  FRDADQDVRQSIERVRTSPFLL 143

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6469> which encodes the amino acid sequence <SEQ ID 6470>. Analysis of this protein sequence reveals the following:

```

      Possible site: 20
      >>> Seems to have no N-terminal signal sequence

50   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2295(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

55 An alignment of the GAS and GBS proteins is shown below.

-2597-

Identities = 109/146 (74%), Positives = 128/146 (87%)

Query: 1 MTTYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60
 + +YF++F+ NQAY LHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN
 5 Sbjct: 1 LMSYFEHFMAANQAYVALHGTAHLPLKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60

Query: 61 AGGRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFINEAFAAQLQRDLGVDMHGHDF 120
 AGGRVT+D++RSLVISQQQ+GTREIVVLHHTDCGAQTFINE FA + LGVD+ G DF
 10 Sbjct: 61 AGGRVTEDMIRSLVISQQQMGTRIVVLHHTDCGAQTFINEGFAKHIHEHLGVDVSGQDF 120

Query: 121 LPFNDIEESVREDVAKLHASPFREE 146
 LPF D+E+SVRED+AK+ AS + ++
 Sbjct: 121 LPFQDVEDSVREDMAKIRASSLISDD 146

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2336

A DNA sequence (GBSx2486) was identified in *S.agalactiae* <SEQ ID 7163> which encodes the amino acid sequence <SEQ ID 7164>. Analysis of this protein sequence reveals the following:

20 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.0932(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG08811 GB:AE004955 phosphoribosylaminoimidazole carboxylase,
 30 catalytic subunit [*Pseudomonas aeruginosa*]
 Identities = 20/27 (74%), Positives = 26/27 (96%)

Query: 1 MFKHAEAEARGRGIKIIAGAGGAAHLP 27
 +F++AEEA GRG+++IIAGAGGAAHLP
 35 Sbjct: 46 LFQYAEAEARGRGLEVIAGAGGAAHLP 72

There is also homology to SEQ ID 910:

Identities = 27/27 (100%), Positives = 27/27 (100%)

40 Query: 1 MFKHAEAEARGRGIKIIAGAGGAAHLP 27
 MFKHAEAEARGRGIKIIAGAGGAAHLP
 Sbjct: 87 MFKHAEAEARGRGIKIIAGAGGAAHLP 113

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2337

A DNA sequence (GBSx2488) was identified in *S.agalactiae* <SEQ ID 7165> which encodes the amino acid sequence <SEQ ID 7166>. Analysis of this protein sequence reveals the following:

50 Possible site: 43
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -6.85	Transmembrane	58 - 74 (53 - 80)
INTEGRAL	Likelihood = -5.79	Transmembrane	103 - 119 (101 - 122)

----- Final Results -----
 55 bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2598-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

There is also homology to SEQ IDs 880 and 9278.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2338

A DNA sequence (GBSx2489) was identified in *S.galactiae* <SEQ ID 7167> which encodes the amino acid sequence <SEQ ID 7168>. This protein is predicted to be short chain alcohol dehydrogenase. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1742 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9357> which encodes amino acid sequence <SEQ ID 9358> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD06605 GB:AE001530 putative oxidoreductase [Helicobacter pylori J99]
Identities = 68/94 (72%), Positives = 79/94 (83%)

Query: 4 IDLLVNNAGLALGLDKSYEADFGDWMINTINNVGLIYLTRCILPKMVEVNRGLIINLGS 63
ID L+NNAGLALGL+K+YE + DW MI+TN+ GL++LTR ILP M+E ++G IINLGS
Sbjct: 76 IDALINNAGLALGLNKAYECELDDEWVMIDTNIKGLLHLRLILPSMIEHDQGTIINLGS 135

Query: 64 XAGTIPYPGANVYGASKAFVKQFSLNLRADLAGT 97
AGT YPG NVYGASKAFVKQFSLNLRADLAGT
Sbjct: 136 IAGTYAYPGGNVYGASKAFVKQFSLNLRADLAGT 169

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7169> which encodes the amino acid sequence <SEQ ID 7170>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9121> which encodes the amino acid sequence <SEQ ID 9122>. Analysis of this protein sequence reveals the following:

Possible site: 12
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.